Gender, Selection into Employment, and the Wage Impact of Immigration

George J. Borjas and Anthony Edo*

June 2024

Abstract

Immigrant supply shocks are expected to reduce the wage of comparable workers. Natives may respond to the lower wage by moving to markets not directly targeted by immigrants and where the wage did not drop. This paper argues that the wage change observed in the targeted market depends not only on how many natives choose to respond, but also on which natives respond. A non-random response alters the composition of the sample of native workers, mechanically changing the average native wage in affected markets and biasing the estimated wage impact of immigration. We document this selection bias in the French labor market, where women accounted for a rapidly increasing share of the foreign-born workforce since 1976. The raw correlations suggest that the immigrant supply shock did not change the wage of native women, but led to a sizable decline in their employment rate. In contrast, immigration had little impact on the employment rate of men, but led to a sizable drop in the male wage. We show that the near-zero correlation between immigration and female wages arises partly because the native women who left the labor force had relatively low wages. Adjusting for the selection bias results in a similar wage elasticity for both native men and women (between -0.7 and -0.9).

JEL classification: *E24, F22, J21, J23*
Keywords: *Immigration, Wages, Selection, Labor Supply, Female Employment*

---

* Borjas: Harvard University and NBER; Edo: CEPII. We are grateful to Christoph Albert, Michael Amior, Axelle Arquié, Gordon Dahl, José De Sousa, Yvonne Giesing, Thomas Grjebine, Daniel Hamermesh, Gordon Hanson, Theodore Joyce, Joan Llull, Joan Monras, Jacques Melitz, Valérie Mignon, Ariell Reshef, Jan Stuhler, Stephen Trejo, Camilo Umana Dajud, Vincent Vicard, and Le Wang for their constructive reactions and assistance. We are also grateful to Christian Dustmann and three anonymous referees for many valuable suggestions.

1. Introduction

All other things equal, an immigration-induced increase in the size of the workforce should reduce the wage of comparable workers. A voluminous literature attempts to estimate the impact of such supply shocks on the wage of native workers. One key insight is that natives may respond by moving to labor markets not directly affected by immigration and where presumably the wage did not drop. For instance, some natives might move to cities that received fewer immigrants and now pay relatively higher wages (Borjas, 2006; Amior, 2020; Monras, 2021), while other natives might leave the labor force altogether (Angrist and Kugler, 2003; Glitz, 2012; Dustmann, Schönberg and Stuhler, 2017).

The native response implies that difference-in-differences comparisons of wages across markets may not identify the wage impact in the market targeted by immigrants. The observed (relative) wage change in the targeted market will reflect both the immediate wage drop after the shock and the potentially attenuating impact of the native response as some of the shock gets transmitted elsewhere.

This insight may be particularly important when examining the differential impact that immigrant supply shocks may have on men and women, two groups that have very different labor supply elasticities at the extensive margin (Blau and Kahn, 2017; Evers, De Mooij and Van Vuuren, 2008). Not surprisingly, the available evidence "does not paint a clear picture of whether women or men are more vulnerable to immigration impacts" (Blau and Mackie, 2016, p. 270).[1]

Specifically, a labor supply response by native workers will likely produce selection biases that differ by gender. For example, male natives may have relatively inelastic supply and the immigrant supply shock would have little impact on the size and composition of the male workforce. In contrast, the labor supply of female natives is much more elastic, and the measured wage impact for native women would then depend not only on the *size* of the native response (i.e., how many women moved in and out of the labor force), but also on its *composition* (i.e., which women moved in and out of the labor force). A sizable, non-random response changes the sample of female native workers, mechanically altering the

[1] A few recent studies examine the potentially different labor market impact of immigration for men and women. These studies include Cortés (2023), Edo and Toubal (2017), Gardner (2020), and Llull (2020).

average native wage in the affected markets. Depending on the context, the selection bias may exacerbate or attenuate the measured impact of immigration on the wage of native women. Although selection bias corrections are rare in the immigration literature, Card (2001, p. 54-56) is one of the first studies to suggest their potential importance. Card performs a back-of-the-envelope calculation that illustrates how the bias can affect occupational wage differentials created by differential supply shocks across occupational groups.

We document the empirical relevance of this selection bias by examining how immigration differentially affected the employment and wages of men and women in the French labor market. The French context is particularly suitable because France experienced a remarkable "feminization" of its immigrant workforce in the past few decades, witnessing a very rapid rise in the female share of foreign workers. If men and women are imperfect substitutes, the increasing number of immigrant women could affect the labor market outcomes of native men and women differently (Acemoglu, Autor and Lyle, 2004; Edo and Toubal, 2017).[2] Female labor supply is also more elastic. As a result, the supply shock may have had a considerable impact on the labor force participation rate of native women, potentially producing a sizable selection bias in the measurement of the wage impact of immigration.

In response to the economic crisis caused by the first oil shock of 1973, the French government stopped recruiting foreign labor in July 1974. In April 1976, however, France granted its foreign-born population the right to family reunification, making it far easier for wives to join their husbands and triggering a rapid rise in the number of female immigrants. Between 1962 and 1975, the immigrant population (aged 18-64) grew by 980.8 thousand persons, and 30.9 percent of this growth was due to female immigration. The immigrant population grew by another 1.1 million persons between 1975 and 2007, and women accounted for 75.6 percent of this increase.

[2] An imperfect degree of substitution between men and women implies that they compete for different types of jobs, so that a rise in the relative supply of women would reduce their relative wage (Topel, 1997). As a result, an immigration-induced increase in the relative supply of female workers would affect wages of native men and women differently.

The top panel of Figure 1 shows how the policy shift led to an immediate drop in the immigrant share of the labor force. In 1975, 10.3 percent of labor force participants were foreign-born. By 1999, the immigrant share had fallen to 8.8 percent. This decline is entirely attributable to a drop in the relative number of immigrant men. In contrast, the immigrant share in the female labor force rose steadily, doubling (from 5.1 to 10.1 percent) between 1962 and 2012. The bottom panel contrasts the French experience with that of the United States. In France, the female share of the foreign-born labor force rose from 21.5 percent in 1962 to 22.8 percent in 1975, and then nearly doubled to 42.4 percent by 1999. In the United States, the female share barely changed between 1970 and 2000, rising only from 39.8 to 41.1 percent over those three decades.

Our analysis is guided by a theoretical framework that illustrates the three key channels through which an immigrant supply shock can change the mean wage of competing workers in a labor market.[3] The first is the wage decline produced by the direct effect of immigration—the downward movement along the labor market's short-run labor demand curve. The second is the attenuation due to a change in the size of the native workforce. Some natives may move out of the market targeted by immigrants, partially reversing the initial shift in the supply curve. The third is the selection bias. Because native workers are not randomly selected from the population, the composition of the native workforce may change after the supply shock, producing a spurious change in the wage.

We use data from population censuses merged with information on labor market outcomes from the Labor Force Surveys (LFS) in the 1982-2016 period. The "raw" data reveal a striking gender asymmetry. The correlation between immigration and wages (across regions and over time) is negative for native men, yet immigration and the male employment rate are uncorrelated. In contrast, the correlation between immigration and female wages is zero, but the correlation between immigration and female employment is strongly negative. We show that the zero wage elasticity for native women is partly an artifact of selection bias. The native women who left the labor market after the supply

[3] There is an additional channel of adjustment as firms increase their capital stock to take advantage of the lower price of labor. Because the empirical analysis reported below does not formally account for this potential response by holding capital constant, our estimates of the wage impact are unlikely to measure short-run effects.

shock tended to be low-wage women, increasing the average wage in the regions targeted by immigrants. After correcting for selection bias, the adjusted wage elasticity for native women is negative and roughly the same size (between -0.7 and -0.9) as that found for native men.

Although our analysis is the first to document how selection biases can produce a gender difference in the observed wage impact of immigration, it is related to other studies that examine the wage and labor supply responses to immigration in various European contexts: Bratsberg and Raaum (2012) for Norway; Dustmann, Schönberg and Stuhler (2017) for Germany; and Ortega and Verdugo (2022) for France. These studies find that low-wage workers are more likely to respond to immigration by leaving or not entering the workforce in the markets targeted by immigrants. The studies "track" the earnings of natives who are continuously employed, thus holding constant the composition of the sample of native workers. The panel reveals a more adverse wage impact than the cross-section data suggest, highlighting the importance of controlling for the composition of the native workforce when measuring the wage impact of immigration.[4]

2. Data and Descriptive Evidence

Our analysis uses data drawn from population censuses and the Labor Force Surveys (LFS) conducted by the French National Institute for Statistics and Economic Studies (INSEE).[5] We use the French censuses from 1962, 1968, 1975, 1982, 1990, 1999, 2007, and 2016 to calculate the size of the population and labor force in each census year (by gender and immigration status). An immigrant is a person born outside France who is either a noncitizen or a naturalized citizen. All other persons are classified as natives.

[4] Note, however, that measuring the impact of immigration by tracking the wage of labor force "survivors" does not necessarily solve the selection problem. The survivors are self-selected from the at-risk population that was exposed to the supply shock, and their experience need not identify the wage impact that would have been observed had the workers who left the labor force remained at work.

[5] The 1962 census extract consist of a random sample of 5 percent of the population, while the post-1962 censuses consist of a random sample of around 10 percent. The high sampling rates allow us to precisely estimate the number of immigrants in different French regions, reducing the role of sampling error in the analysis. The LFS surveys a random sample of the French population, with a sampling rate of 0.3 percent before 2002, and between 0.7 percent and 1.0 percent after 2002. We use the personal weight throughout the analysis to make our LFS samples representative of the French population.

The annual LFS reports a worker's monthly wage net of employee payroll tax contribution beginning in 1982.[6] The LFS also reports a person's labor force and employment status during the reference week, and many demographic characteristics (including age, gender, nationality, education, marital status, and number of children). Our analysis focuses on the monthly real wage of full-time native workers to have a more precise measure of the "price of labor." The LFS is designed to be representative of the population at the regional level (there are 22 regions in European France), and we conduct our empirical analysis mainly at this geographic level.

Our sample is restricted to persons aged 18-64 living in European France. We exclude all persons who are self-employed, are in military occupations, are enrolled in school, or do not report their educational attainment. We also use the reported average number of hours worked per week in the LFS to exclude from our wage analysis observations that have extreme values of the hourly wage. Specifically, we exclude workers who are in the top 0.5% or bottom 0.5% of the hourly wage distribution.

Table 1 summarizes the data. There was a sizable increase in the employment rate of native women between 1962 and 2016, almost doubling from 37.1 to 70.1 percent. At the same time, the employment rate of native men declined, from 89.3 to 73.6 percent. The immigrant supply shock was roughly similar for low- and high-educated native women. The immigrant share rose from 2.8 to 9.2 percent for women with a baccalaureate degree (i.e., at least a high school education) and from 5.4 to 14.1 percent for women without the degree.[7] In contrast, immigration had a larger impact on the number of high-educated men.

We merge the employment rates and the data on the number of immigrants reported in the censuses with the concurrent LFS wage data for native workers. The merged data helps illustrate the raw relationship between immigration and native labor market outcomes across French regions over the 1982-2016 period. In this descriptive analysis, the unit of observation is a region-year cell. For each cell, we estimated the mean

[6] The 1982 LFS does not provide information on a person's nationality at birth. We use the sample of persons born in France to compute the monthly wage of natives in 1982.

[7] A baccalaureate degree is granted to persons who completed high school by passing a French exam named the "Baccalauréat," which gives access to college or an equivalent diploma. Table 1 shows that only 25.4 percent of the native labor force in 1982 had a Baccalaureate degree; by 2016, this fraction had increased to 60.9 percent.

log monthly wage of full-time workers (separately by gender) as well as the immigrant share defined by $m_{rt} = \log(1 + M_{rt}/N_{rt})$, where M_{rt} gives the total number of (male and female) immigrants in the labor force in region r at time t and N_{rt} gives the corresponding number of natives.[8] We calculated the adjusted mean wage as the residual from a regression (estimated separately by gender) of the mean log monthly wage on vectors of region and year fixed effects. We also calculated the adjusted supply shock by obtaining the residuals from a regression of m_{rt} on vectors of region and year fixed effects.

Figures 2A and 2B document the gender asymmetry in the relation between immigration and wages in France.[9] The figures show a weak positive correlation between immigration and female wages (the regression coefficient is 0.19, with a standard error of 0.12), but a strong negative correlation between immigration and male wages (the coefficient and standard error are -0.31 and 0.15, respectively). Using a similar approach, we calculated the gender-specific adjusted employment rates for each region-year cell. These data further illustrate the gender asymmetry. Figures 2C and 2D show a strong negative correlation between employment and immigration for native women (the coefficient and standard error are -0.98 and 0.07), and a zero correlation for native men (the coefficient and standard error are -0.02 and 0.06).[10]

The data thus suggest that immigration affected the labor market opportunities of native men along the wage margin and of native women along the employment margin. These correlations suggest that immigration may have had a crowd-out effect on female employment, attenuating the (initial) wage reduction produced by the supply shock. The attenuation would be magnified if the women who left the labor market had relatively low

[8] Bratsberg and Raaum (2012) and Edo and Özgüzel (2023) also use this definition of the immigrant share. Most studies in the literature define the supply shock as either M/N or as $M/(M + N)$. We show below that our definition of m_{rt} is the measure of the supply shock implied by a labor demand model. Our results would be very similar if we used the approximations in the literature.

[9] The asymmetric impact of immigration on the employment of native men and women is also reported by Angrist and Kugler (2003) in Europe, Edo (2020) in France, and Gardner (2020) in the United States.

[10] Appendix A reports robustness checks that further document the asymmetry illustrated in Figure 2. Figure A shows that the negative relationship between the employment rate and immigration is strong and significant regardless of whether native women are full- or part-time workers. Table A1 shows that the descriptive correlations summarized in Figure 2 are robust if we begin the sample period in 1968, instrument the share of immigrants (by using a shift-share instrument), and use 94 departments (instead of 22 regions) to define the local labor market.

wages. In other words, the zero correlation between wages and immigration for native women may simply be a consequence of elastic female labor supply—and the ensuing selection bias—and may not necessarily reflect the initial wage impact of the supply shock.

3. Theoretical Framework

3.1. Selection Bias and the Wage Impact of Immigration

Consider a stylized two-period model summarized by:

Wage offer at $t = 0$: $$\log w_{ik0} = \mu_k + \epsilon_{i0}, \quad (1a)$$

Wage offer at $t = 1$: $$\log w_{ik1} = \mu_k + \delta_k + \epsilon_{i1}, \quad (1b)$$

Reservation wage: $$\log \mathcal{R}_i = \bar{\mathcal{R}} + h_i, \quad (1c)$$

where w_{ikt} gives the wage of person i in labor market k at time t; μ_k is the initial mean of the population wage distribution; $\mathcal{R}_i$ is the reservation wage; and $\bar{\mathcal{R}}$ is the mean (log) reservation wage. [11] The ϵ's and h capture (unobserved) individual variation in wage offers and reservation wages.

The parameter δ_k measures the wage impact of an immigrant supply shock that hits market k between the two periods. We assume that this supply shock only shifts the mean of the population wage distribution. To fix ideas, our discussion focuses on the case where immigrants and natives are substitutes, so that $\delta_k < 0$ (although the selection problem also arises when immigrants and natives are complements).

Figure 3 illustrates the selection bias if we assume that a single (unobserved) "skills" factor, ω, determines both the wage offer and the reservation wage (i.e., $\epsilon_{i0} = \beta_w\ \omega_i$; $\epsilon_{1i} = \beta_w\ \omega_i$; and $h_i = \beta_h\ \omega_i$). This assumption implies that $Corr(\epsilon_0, h) = Corr(\epsilon_1, h) = Corr(\epsilon_0, \epsilon_1) = 1$. As drawn, the wage curves indicate that the returns to skills are larger in the labor market than in household production, leading to a positively selected workforce.

---

[11] Depending on the context, a labor market k may be viewed as a region, a skill group, or a region-skill group.

All persons with skills above threshold θ_0 work at t = 0 and the supply shock increases this threshold from θ_0 to θ_1, so that the labor force participation rate falls.[12] Suppose that the distribution of ω is uniformly distributed over the interval depicted in the figure. The mean wage of *workers* in the initial period is given by point A, the midpoint of the wage curve between θ_0 and the maximum wage. Similarly, the mean wage of workers after the supply shock is given by point B (the midpoint between θ_1 and the maximum wage). The vertical difference between A and B (which is very small) does not identify the wage impact δ_k (i.e., the vertical difference between the two wage curves).

The change in the average wage earned by native workers depends crucially on *how many* and *which* native workers respond to the shock. The exit of low-skill natives from the labor force artificially increases the average wage, so that a comparison of mean wages across cross-sections could end up suggesting that immigration had little impact or even increased wages. In short, the wage change observed between cross-sections (i.e., the classic identification strategy in the literature) does not identify how a supply shock shifts the mean of the wage distribution.

This insight suggests that we could retrieve the correct wage effect δ_k by applying a selection correction to earnings functions estimated in *each* cross-section. A selection-corrected wage equation in the initial cross-section identifies the mean wage in the population at t = 0 (or point A_S, the midpoint of the wage curve). Similarly, a selection-corrected wage equation in the second cross-section estimates B_S, the mean wage in the population after the shock. The vertical difference between A_S and B_S identifies δ_k.

It is instructive to derive the selection bias produced by the cross-section estimator in the canonical case where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and $h_i \sim N(0, \sigma_h^2)$. We continue to assume that the supply shock only shifts the mean of the population wage distribution by δ_k. The event I_{ikt} indicating if person i in market k works at time t is defined by:

$$I_{ik0}: \quad v_{i0} = \epsilon_{i0} - h_i > \bar{\mathcal{R}} - \mu_k = \theta_{k0}. \tag{2a}$$

$$I_{ik1}: \quad v_{i1} = \epsilon_{i1} - h_i > \bar{\mathcal{R}} - \mu_k - \delta_k = \theta_{k1}, \tag{2b}$$

---

[12] The exit of some natives from the labor force implies that part of the wage drop observed immediately after the supply shock is attenuated. The parameter δ_k then measures the "net" impact of the shock. The attenuation effect is discussed in greater detail below.

where $v_{it} \sim N(0, \sigma_v^2)$.[13] The labor force participation rate in market k at time t is $\pi_{kt} = 1 - \Phi(\theta_{kt}/\sigma_v)$, where $\Phi(\cdot)$ represents the standard normal distribution function.

The average wage change observed across cross-sections is defined by:

$$\Delta \log w_k = E[\log w_{ik1} \mid I_{ik1}] - E[\log w_{ik0} \mid I_{ik0}] = \delta_k + E[\epsilon_{i1} \mid I_{ik1}] - E[\epsilon_{i0} \mid I_{ik0}]\,. \qquad (3)$$

Using standard results from the selection literature (Gronau, 1974; Heckman, 1979) yields:

$$E[\epsilon_{it} \mid I_{ikt}] = \sigma_\epsilon\, \rho_{\epsilon v}\, \lambda(\pi_{kt})\,, \qquad (4)$$

where $\rho_{\epsilon v} = Corr(\epsilon_{it}, v_{it})$; and $\lambda(\pi_{kt}) = \phi(\theta_{kt})/\pi_{kt}$, with $\phi(\cdot)$ representing the standard normal density. We can then rewrite equation (3) as:

$$\Delta \log w_k = \delta_k + \sigma_\epsilon\, \rho_{\epsilon v}\, [\lambda(\pi_{k1}) - \lambda(\pi_{k0})]. \qquad (5)$$

If immigration has an adverse wage impact ($\delta_k < 0$), the participation rate of natives falls ($\pi_{k1} < \pi_{k0}$) and $[\lambda(\pi_{k1}) - \lambda(\pi_{k0})] > 0$.[14] Equation (5) then implies that the wage impact estimated from repeated cross-sections is "too positive" if the workforce is positively selected from the population ($\rho_{\epsilon v} > 0$) and "too negative" if the workforce is negatively selected ($\rho_{\epsilon v} < 0$). More generally, if the native workforce is not randomly selected ($\rho_{\epsilon v} \neq 0$), equation (5) implies that selection bias exists whenever $\delta_k \neq 0$ (so that $\pi_{k1} \neq \pi_{k0}$). In other words, the cross-section estimator yields a biased measure of the wage impact of immigration regardless of whether immigrants and natives are substitutes or complements.

As noted above, some recent studies propose the alternative strategy of tracking the persons who worked both before and after a supply shock to identify δ_k. In fact, the special case of the model illustrated in Figure 3 indicates that the panel strategy identifies the wage impact *without* using any selection correction. The observed wage in the panel of

---

[13] The assumption that immigration only changes the mean of the wage distribution implies that $Var(\epsilon_{i0}) = Var(\epsilon_{i1}) = \sigma_\epsilon^2$. It follows that $\sigma_{v_0}^2 = \sigma_{v_1}^2 = \sigma_v^2$ and that $Corr(\epsilon_{i0}, v_{i0}) = Corr(\epsilon_{i1}, v_{i1}) = \rho_{\epsilon v}$.

[14] The inverse Mills ratio is a negative function of the participation rate (Heckman, 1979, p. 156), so that $\lambda(\pi_{k1}) > \lambda(\pi_{k0})$ when $\pi_{k1} < \pi_{k0}$.

workers who are continuously employed is the midpoint of the wage curves after the threshold θ_1 in each cross-section, or points A_P and B. The difference $(A_P - B)$ equals δ_k. As Borjas and Edo (2021) show, however, this property of panel data does not necessarily generalize when the random variables in $(1a)$-$(1c)$ have a joint normal distribution. In that canonical case, the panel estimator identifies δ_k only if $Corr(\epsilon_{i0}, \epsilon_{i1}) = 1$, so that individual earnings are perfectly correlated before and after the supply shock.

In sum, the self-selection of the native workforce can bias standard estimates of the wage impact of immigration if: (a) native workers are not randomly selected from the population; *and* (b) supply shocks affect the labor force participation decision of natives. The identification of the wage impact will then require either a selectivity-corrected analysis of the mean wage of workers across repeated cross-sections, or a selectivity-corrected analysis of the wage growth observed in a panel of persons who worked continuously through the sample period.

The cross-section approach, which we pursue below, has two advantages. First, panel data suitable for the study of immigrant supply shocks over two or three decades are limited in the French context. Second, the cross-section selection correction is a straightforward application of the Heckman two-step procedure, applied to each cross-section to retrieve the population mean wage in a particular market at a point in time. Although some studies have used panel data to correct for selection bias (e.g., Bratsberg and Raaum, 2012), the panel sample is "double-selected", composed of persons who worked prior to the supply shock and then filtered further to include only the labor force survivors.

3.2. Empirical Specification of the Wage Impact

We modeled the wage impact of immigration as a shift in the mean of the population wage distribution, but left open the question of how the shifter δ_k should be specified in a regression framework. Consider a labor market represented by the Cobb-Douglas production function:

$$Q_{kt} = A_t\, K_{kt}^{\eta}\, L_{kt}^{1-\eta}, \tag{6}$$

where Q_{kt} denotes output in market k at time t, and K_{kt} and L_{kt} denote capital and labor, respectively. The marginal productivity condition implies that the market wage w_{kt} is:

$$\log w_{kt} = \varphi_{kt} - \eta \log L_{kt} \,, \tag{7}$$

where $\varphi_{kt} = \log\left[(1-\eta)A_t K_{kt}^{\eta}\right]$, a parameter specific to cell (k, t). The coefficient η gives the wage elasticity, which equals capital's share of income in a Cobb-Douglas framework.

This market has received immigrant supply shocks in the past, and the workforce has N_{k0} natives and M_{k0} immigrants in the base period. Suppose that immigrants and natives are perfect substitutes. It is convenient to rewrite equation (7) as:

$$\log w_{k0} = \varphi_{k0} - \eta \log(M_{k0} + N_{k0}) = \varphi_{k0} - \eta\, m_{k0} - \eta \log N_{k0} \,, \tag{8}$$

where the immigrant share $m_{kt} = \log(1 + M_{kt}/\, N_{kt})$ approximates the fraction of the workforce that is foreign-born (as a fraction of native workers in the same period).

A new influx of immigrants enters the market, increasing the total number of foreign workers to M_{k1}. Consider the short-run impact of this supply shock (i.e., holding capital constant). Suppose that immigrant labor supply is perfectly inelastic. Native labor supply, however, might respond to the supply shock, so that the total number of native workers changes to N_{k1}. The native response consists exclusively of movements in and out of the labor force (and not migration to labor market k'). The market wage *after* natives have responded is:[15]

$$\log w_{k1} = \varphi_{k1} - \eta \log(M_{k1} + N_{k1}) = \varphi_{k1} - \eta\, m_{k1} - \eta \log N_{k1} \,. \tag{9}$$

There will be some within-market wage dispersion because persons in market k (though they share characteristics that help define the market, such as location) exhibit some differences. Natives may differ in their motivation for work or have an ethnic background that is favored or penalized by employers. The wage offer made by firms to

[15] Borjas and Edo (2021, p. 18) show that the assumption of imperfect substitution between immigrants and natives in a nested CES framework produces an equation analogous to (10), but changes the interpretation of the coefficients as they would also involve the elasticity of substitution between the two groups. It would also imply that the coefficients of the two regressors in equation (10) need not be equal.

potential native workers then depends not only on market conditions, but also on (unobserved) individual characteristics captured by ϵ_{it}:

$$\log w_{ikt} = \varphi_{kt} - \eta\, m_{kt} - \eta \log N_{kt} + \epsilon_{it}, \tag{10}$$

where $\epsilon_{it} \backsim N(0, \sigma_\epsilon^2)$.

Combining equation (10) with the selection bias expression in equation (5), the *observed* change in the mean wage of native workers in repeated cross-sections is:

$$\begin{aligned} \Delta \log w_k &= E[\log w_{ik1} \mid I_{ik1}] - E[\log w_{ik0} \mid I_{ik0}] \\ &= \varphi - \eta\, \Delta m_k - \eta\, \Delta \log N_k + \sigma_\epsilon\, \rho_{v\epsilon}\, [\lambda(\pi_{k1}) - \lambda(\pi_{k0})]. \end{aligned} \tag{11}$$

Equation (11) shows that immigration has three distinct effects on the wage change observed in market k. The first is the direct short-run effect of the shock Δm_k, captured by the (negative) wage elasticity η. This is the downward movement along the short-run labor demand curve in the absence of any native response. The second captures the possibility that immigrants crowd out the supply of natives. The percent change in the number of native workers generates its own attenuating effect, as that supply response moves the labor market back up the labor demand curve. The third gives the selection bias.

The regression equation typically used in the literature differs from (11) in two distinct ways. First, it ignores the selection bias correction. Second, it ignores the variable measuring the size of the native workforce. By excluding $\Delta \log N_k$, the generic regression is estimating a type of reduced-form effect. Following Borjas and Monras (2017), we can write the change in the number of native workers as a function of the supply shock:

$$\Delta \log N_k = \gamma\, \Delta m_k, \tag{12}$$

where γ $(-1 \leq \gamma \leq 0)$ captures the crowd-out effect, approximating the number of native workers who leave the labor market for every immigrant who enters. Substituting equation (12) into (11) yields a type of "reduced-form" labor demand function:

$$\Delta \log w_k = \varphi - \eta(1+\gamma)\, \Delta m_k + \sigma_\epsilon\, \rho_{v\epsilon}\, [\lambda(\pi_{k1}) - \lambda(\pi_{k0})]. \tag{13}$$

Abstracting from the selection bias correction, equation (13) is the regression typically estimated in the literature. This regression produces an estimate of $-\eta(1+\gamma)$, the "reduced-form" wage elasticity that incorporates the native supply response.[16]

4. Econometric Framework

4.1. The Wage Equation

We estimate the selection-adjusted wage impact of immigration by turning to individual-level data in a pooled sample of cross-sections and applying the Heckman selection correction. Consider the earnings function:

$$\log w_{irt} = \theta_{at} + \theta_{et} + \alpha P_{it} + \varphi \lambda_{it} + \theta_{rt} + \mu_{it}, \qquad (14a)$$

where $\log w_{irt}$ gives the log monthly wage of native worker i in region r at time t; θ_{at} and θ_{et} are vectors of interacted age-time and education-time fixed effects, respectively; P_{it} is a vector of other personal characteristics; λ_{it} is the inverse Mills ratio calculated from a first-stage probit on the probability that the individual is employed (discussed below); and θ_{rt} is a vector of interacted region-time fixed effects.[17]

If the inverse Mills ratio were excluded from the regression in (14a), the elements of the vector $\hat{\theta}_{rt}$ give the (age- and education-adjusted) mean wage of *workers* in cell (r, t). These fixed effects are biased estimates of the mean of the *population* wage distribution in the cell, as they are calculated using a sample of working natives that varies non-randomly across regions and over time. If the standard assumptions of the Heckman selection correction hold, the inclusion of the inverse Mills ratio in equation (14a) implies that $\hat{\theta}_{rt}$

---

[16] Appendix B shows that regressing the gender-specific wage change on the *total* immigrant supply shock, as implied by equation (13), does not identify the "reduced-form" wage elasticity when men and women are imperfect substitutes. In the more general case, the measured wage effect will also depend on the difference in supply elasticities between men and women, the elasticity of substitution between the two groups, the share of women or men in the workforce, and the relationship between the total supply shock and the shock in the female or male labor market. The Appendix also shows that the supply shock has a larger impact on the gender wage gap the smaller the elasticity of substitution between men and women.

[17] The age fixed effects consist of six age categories (18-24, 25-32, 33-39, 40-47, 48-55, 56-64) and the education fixed effects consist of four education categories (persons with less than a high school diploma, high school graduates, persons with some college, and college graduates).

consistently estimates the (adjusted) mean wage of the population in cell (r, t). We estimate this individual-level earnings functions separately in the samples of working men and women.

The framework developed in the previous section showed that the wage impact of immigration can be identified by examining how supply shocks shift the mean of the population earnings distribution across markets in repeated cross-sections. We thus estimate the following second-stage regression using cell-level data:

$$\hat{\theta}_{rt} = \theta_r + \theta_t + \alpha_M \, m_{rt} + \alpha_N \log N_{rt} + \xi_{it}, \tag{14b}$$

where θ_r and θ_t are vectors of region and time fixed effects, respectively.[18]

Following Dustmann, Schönberg, and Stuhler (2017) and Jaeger, Ruist, and Stuhler (2018), we define the immigrant share m_{rt} at the region-year level (instead of assigning workers to different skill groups and calculating a supply shock specific to a region-skill-year cell). This strategy accounts for all channels through which a supply shock in region r can affect the wage of workers in that region. Put differently, the estimate of α_M captures the sum of the "own" effect of a specific supply shock on the wage of competing workers, the complementary effects on the wage of workers with different skills, and the wage adjustments produced by changes in capital accumulation. Moreover, this approach does not pre-assign workers to specific skill groups, avoiding the mismeasurement introduced by the possibility that employers might downgrade the skills that immigrants offer to the labor market (Dustmann, Frattini, and Preston, 2013).

Our empirical framework only corrects for the selection bias produced by the endogenous native labor supply decision. There are other native responses that also produce a self-selected sample and contaminate estimates of the wage impact of immigration. For example, equation (14b) may give a biased estimate of the wage impact if natives respond to local supply shocks by moving to other areas (Borjas, 2006; Dustmann,

[18] By combining equations (14a) and (14b), the two-stage model collapses into a single regression:

$$\log w_{irt} = \theta_{at} + \theta_{et} + \theta_r + \theta_t + \alpha_M m_{rt} + \alpha_N \log N_{rt} + \varphi \lambda_{it} + \mu'_{it}.$$

We prefer the two-equation framework as it more clearly shows the identification of the wage impact and makes our regression analysis comparable to the cell-level studies that dominate the immigration literature.

Fabbri and Preston, 2005; Edo, 2019). This bias, however, is unlikely to affect the analysis in the French context because we define the local labor market at a *regional* level. The available evidence suggests that native regional migration was not correlated with immigrant supply shocks in the post-1960 period (Edo, Giesing, Poutvaara and Öztunc, 2019; and Edo, 2020).[19]

The three key variables in the regression model in equations ($14a$) and ($14b$) need to either be estimated (the inverse Mills ratio λ_{it}) or are endogenous (m_{rt} and $\log N_{rt}$). We now turn to a discussion of the first stage probit depicting an individual's labor force participation decision and of the instruments used to correct for the endogeneity.

4.2. The Inverse Mills Ratio

We construct the inverse Mills ratio by first estimating a probit model that relates a native person's work decision to the various regressors in the model, including a vector of "instruments" Z that, by assumption, do not enter the wage equation:

$$Pr(EMP_{irt} = 1) = \Phi(\theta_{at} + \theta_{et} + \alpha_P P_{it} + \alpha_Z Z_{it} + \theta_{rt} + v_{it}). \tag{15}$$

Note that the probit equation does not include any immigration-related variables, but instead includes the vector of region-time fixed effects θ_{rt} (which subsumes all potential measures of the local shock and native supply response). We categorize the population into working or not working based on person i's employment status in the reference week of the LFS data. In other words, EMP_{irt} is a binary variable indicating whether native person i in region r at time t is employed. We estimate the probit model in equation (15) separately for men and women.

Because there are gender differences in the determinants of labor supply and wages, we use slightly different baseline specifications of the wage and probit regressions in equations ($14a$) and (15) for the two groups. Our approach for the analysis of female outcomes follows the literature (Mulligan and Rubinstein, 2008; Blau and Kahn, 2017, p.

[19] Using our sample and IV strategy, we estimated a regression relating the rate of change in the adult native population in each region on the respective rate of change in the immigrant population (Appendix Table A2). The IV coefficient was insignificant, showing that immigration did not affect the mobility of natives across French regions in the 1982-2016 sample period.

810; Machado, 2017; Maasoumi and Wang, 2019). The probit includes variables that adjust for differences in the female reservation wage, and these differences are captured by marital status and the presence of young children (under age 6) in the household.[20] It is typically assumed that these family characteristics affect the reservation wage of women, but do not affect their wage.

The LFS data allow us to expand this generic specification as it contains a rough measure of household wealth (so that we can also control for income effects on labor force participation). The available measure of household wealth indicates if the person owns their home free of any debt.[21] As long as leisure is a normal good, higher levels of household wealth increase the reservation wage and should have a negative effect on the probability of participating in the labor force.

The regression specification for the joint study of female employment and wages can then be summarized as follows: The individual-level wage regression will include vectors of age-time, education-time, and region-time fixed effects. The probit regression includes all these variables *plus* the family characteristics (marital status and the presence of young children) and household wealth. The independent variation in the inverse Mills ratio in the female wage regression is generated by the presence of both family characteristics and household wealth in the first-stage probit.

The U.S. labor supply literature often asserts that the selection problem is not empirically relevant for men (Pencavel, 1986, p. 55; Mulligan and Rubinstein, 2008). This assumption, however, may not be applicable in France, where the unemployment rate for prime-age men is high, and the assumption that male workers are randomly selected from

[20] The marital status variable in the LFS classifies individuals into one of four groups: single, widowed, divorced, or married. We pool all single, divorced, or widowed natives into the "unmarried" group.

[21] The fraction of persons who own their home without a mortgage rose from 22.0 to 32.2 percent between 1982 and 2016. The homeownership information was not collected for a random half of the sample in the 2016 LFS cross-section. We impute the missing values by running a probit regression in the pooled 1982-2016 data that relates the homeownership indicator (if available) to age, education, interacted region-time fixed effects, and a full set of interactions between gender, marital status, presence of young children, and region fixed effects. We impute a value of 1 or 0 to the missing observations based on whether the predicted probability of home ownership was above or below 0.6. Our results are similar if we simply exclude the 2016 observations that had missing information on homeownership assets.

the population is less plausible. During our sample period, for example, the average unemployment rate of native men aged 25-59 was 8.1 percent.

Our specification of the regression models for the joint study of male labor supply and earnings differs slightly from what is typically used in the female context because marriage may have a productivity-related positive effect on male earnings (Choi, Joesch, and Lundberg, 2008; and McDonald, 2020), and fatherhood may also increase male earnings (Lundberg and Rose, 2000). In other words, even if these family characteristics did not affect the male reservation wage, they would need to enter *both* the probit and the individual-level wage regressions because they affect the male wage directly. As a result, the family variables do not produce independent variation for the inverse Mills ratio in the male wage regression. This independent variation is instead produced by the measure of household wealth that we assume only affects reservation wages.

The baseline regression specification for the study of male employment and wages is: The individual-level wage regression includes vectors of age-time, education-time, and region-time fixed effects, as well as marital status and presence of young children. The probit regression includes all these variables *plus* the measure of household wealth.

We report below two alternative empirical exercises to evaluate the robustness of our selectivity-corrected estimates. First, we show that our estimates of the wage elasticity are robust to alternative specifications of the selection model (including the assumption of no selection for men). Second, we isolate a subsample of female native workers for whom selection into employment is unlikely to matter (specifically, young single women without children) and show that the (uncorrected) wage elasticity for this subsample is very similar to the selectivity-corrected wage elasticity for the entire sample of women. In the spirit of the identification-at-infinity method of correcting for selection bias (Chamberlain, 1986; Heckman, 1990; Mulligan and Rubinstein, 2008; and Blau, Kahn, Boboshko and Comey, 2021), this approach does not require the specification of identifying variables in a first-stage probit regression.

4.3. Endogeneity of the Immigrant Supply Shock

It is well known that estimating the cell-level model in (14*b*) using OLS produces inconsistent estimates of the wage impact of immigration because of the endogenous sorting of immigrants across regions. To address this issue, we use an instrumental variable approach, with the instrument based on past immigration patterns (Altonji and Card, 1991).

To build our instrument, we use the 1968 spatial distribution of immigrants from a given nationality for a given education group to predict the sorting of immigrants in subsequent periods (Edo, Giesing, Poutvaara and Öztunc, 2019). We use 11 nationality groups and four education groups.[22] We predict the number of immigrants for each region-time cell at time t ($t > 1968$) by multiplying the 1968 spatial distribution of immigrants in each origin-education group by the total number of immigrants from that group at time t:

$$\widehat{M}_{rt} = \sum_n \sum_e \frac{M_r^{ne}(1968)}{M^{ne}(1968)} \cdot M_t^{ne}, \tag{16}$$

where M_{rt}^{ne} gives the number of immigrants in year t in national origin group n, education group e, and region r; and $M_t^{ne} = \sum_r M_{rt}^{ne}$. We use an analogous approach to predict the number of natives in the region because the actual number of natives is unlikely to be independent from regional conditions:

$$\widehat{N}_{rt} = \sum_e \frac{N_r^e(1968)}{N^e(1968)} \cdot N_t^e. \tag{17}$$

The shift-share instrument is then defined by:

$$\widehat{m}_{rt} = \log\left(1 + \frac{\widehat{M}_{rt}}{\widehat{N}_{rt}}\right). \tag{18}$$

Despite their widespread use, shift-share instruments may not satisfy the exclusion restriction required by the IV strategy. As Goldsmith-Pinkham, Sorkin, and Swift (2020, p. 2593) note: "The identification concern is whether [past local immigrant shares are] correlated with changes in the outcome, and not levels of the outcome." [23] Such a correlation between past immigration to a particular region and current wage growth may

[22] The nationality groups are: Italian, Portuguese, Spanish, other European, Algerian, Moroccan, Tunisian, other African, Turkish, the rest of the world, and French for those immigrants who acquired the French citizenship. The education groups are persons with less than a high school diploma, high school graduates, persons with some college, and college graduates.

[23] This point applies precisely to the typical empirical framework in the immigration literature that either estimates a wage level equation using repeated cross sections *and* includes unit fixed effects, or specifies the regression equation in first-differences.

arise if: *(a)* local economic conditions that influenced past immigrant settlement patterns are serially correlated over time (Dustmann, Fabbri and Preston, 2005); and/or *(b)* current economic outcomes are still adjusting to past immigration (Jaeger, Ruist and Stuhler, 2018). In short, the shift-share instrument in equation (18) is valid only if the 1968 spatial distributions of immigrants and natives are uncorrelated with the unobserved component of regional wage growth after the 1980s.

Although there is no formal way of testing for the exogeneity of an instrument, the results reported in Appendix C, which follow the discussion in Goldsmith-Pinkham, Sorkin and Swift (2020), show that the exogeneity assumption is likely to be satisfied in our context. Specifically, we constructed the Rotemberg weights for each origin-specific share of immigrants used in the construction of the Bartik-type instrument. Appendix Table C1 reports the correlation between the regional origin-specific immigrant shares and initial period characteristics in 1968. There is no significant relationship (with only one exception) between these origin shares and the regional wage of natives and immigrants. We also examined if the pre-1982 regional change in the wage and employment of natives was correlated with the predicted immigrant shares over the 1982-2016 period. The estimated coefficients are not statistically significant, suggesting that the pre-existing trends in native labor market opportunities and the predicted regional evolution of immigrant shares are not directly linked.

4.4. Endogeneity of Native Labor Supply

Although the generic regression model used in the immigration literature simply relates the wage in a particular market to the immigrant share in that market, the labor demand framework indicates that a fully specified regression model should also include the size of the native labor force. Few studies, however, pursue this implication of the theory (exceptions include Borjas, 2003; and Bratsberg, Raaum, Røed and Schøne, 2014). As shown in Section 3, the exclusion of this variable identifies a reduced-form estimate of the wage elasticity that is contaminated by the size of the crowd-out effect.

The size of the native labor force is endogenous to local economic conditions. Our instrument combines the shift-share projection of the native population with information on gender and such (presumed) exogenous variables as the presence of young children in the household. The summary statistics in Table 1 suggest that a major determinant of changes in the size of the native workforce was the increase in the employment rate of women. As in other countries, the presence of young children deters female labor supply in France (Piketty, 1998; Gurgand and Margolis, 2008). Let ψ_{rt} be the fraction of the native population in region r at time t that is female and that does *not* have children under the age of 6.[24] Our instrument for the (log) size of the native workforce is given by:

$$\log \hat{F}_{rt} = \log \left[\psi_{rt} \cdot \hat{\mathcal{N}}_{rt}\right], \tag{19}$$

where $\hat{\mathcal{N}}_{rt}$ is an adjusted measure of the shift-share prediction $\hat{N}_{rt}$ of the native population. The variable $\hat{F}_{rt}$ thus gives the predicted female native labor force in region *r* at time *t*.

The construction of $\hat{N}_{rt}$ in equation (17) only took into account the geographic allocation of natives at the time of the 1968 cross-section, and ignored region-specific long-run trends that were systematically changing that allocation prior to our sample period. Unlike changes in the population of immigrants, where sizable shocks can occur due to exogenous policy shifts or economic and political shocks in source countries, future projections of the native population are much more dependent on pre-existing trends.

To construct the instrument in (19), we adjust the shift-share projection $\hat{N}_{rt}$ for the long-term regional differences in population growth rates. We calculate the (baseline) annual growth rate of the native population in region r between 1968 and 1982, g_r, as well as the growth rate of the shift-share projection over the same period, $\hat{g}_r$, and define $\Delta g_r = g_r - \hat{g}_r$. The adjusted shift-share projection is then given by:

$$\hat{\mathcal{N}}_{rt} = \hat{N}_{rt}(1 + \Delta g_r)^{t-1968}. \tag{20}$$

The adjusted projection $\hat{\mathcal{N}}_{rt}$ equals the "cross-section" projection $\hat{N}_{rt}$ if the geographic allocation of natives has been constant prior to the sample period (i.e., $\Delta g_r = 0$).[25]

---

[24] The variable ψ_{rt} equals the share of the population that is female (drawn from the census) times the share of the female population that does not have young children (drawn from the LFS).

[25] The population data for the Île-de-France region, which includes Paris, illustrates the importance of this adjustment. This region's population grew by only 0.5 percent per year between 1968 and 1982, as compared to a national growth rate of 1.3 percent. The 2016 shift-share prediction $\hat{N}_{rt}$ for Île-de-France is 8.5 million persons, as compared to an actual native population of only 5.0 million. The adjustment in equation (20) produces a prediction of 4.5 million.

The exclusion of the log N_{rt} variable from the typical regression model in the literature is undoubtedly due to the difficulty in finding good instruments for native labor supply. Because of the absence of compelling exogenous shocks in native labor supply, our empirical analysis will report estimates of the wage impact of immigration both excluding and including the native labor supply variable. The evidence will demonstrate that the key insight of our framework—i.e., that selection biases matter when estimating the wage impact—is valid even when we restrict our attention to the "reduced form" wage elasticity identified by the generic equation in the literature.

5. Empirical Results

5.1. First-Stage IV Estimates

Table 2 presents the first stage of our baseline IV wage regressions for both native women and native men. Our simplest regression specification relates the wage to the immigrant share. Panel A of the table presents the first-stage regression associated with this model, where we regress m_{rt} (i.e., the single endogenous regressor) on $\widehat{m}_{rt}$ (i.e., the shift-share instrument defined in equation (18)), region, and time fixed effects.

Not surprisingly, the first stage shows a strong positive and significant correlation between the instrument and the endogenous variable. We also report the Kleibergen-Paap rk Wald F statistics as this test accounts for the non-i.i.d. structure of the residual (Kleibergen and Paap, 2006). They are larger than the lower bound of 10 suggested by the literature on weak instruments (Stock, Wright and Yogo, 2002), indicating that our IV estimates are unlikely to suffer from a weak instrument problem.

Panel B reports the first-stage estimates for the expanded specification that has two endogenous variables, the immigrant share m_{rt} and native labor supply (log N_{rt}). The instruments are the log predicted population of immigrants $\widehat{M}_{rt}$ defined in equation (16), and the log predicted female native labor force $\widehat{F}_{rt}$ defined in equation (19).[26] All regressions again include region and time fixed effects.

[26] We use log $\widehat{M}_{rt}$ instead of $\widehat{m}_{rt}$ as an instrument to avoid potential collinearity issues arising from the fact that $\widehat{m}_{rt}$ and log $\widehat{F}_{rt}$ are both functions of the shift-share prediction of the native population.

The results indicate that the immigrant share is positively correlated with $log\widehat{M}_{rt}$ and negatively correlated with $log\widehat{F}_{rt}$. The positive correlation is in line with the literature on the immigrant shift-share instrument, while the negative correlation probably arises because a rise in the predicted female native labor force would mechanically reduce the ratio of immigrant to native workers. There is also a very strong positive correlation between the predicted female native labor force and the size of the native labor force.

To evaluate the strength of our two instruments, we use the IV first-stage F-statistics for the case of multiple endogenous variables proposed by Sanderson and Windmeijer (2016). The first-stage F-tests of excluded instruments are between 12.9 and 16.5, indicating that our instruments are reasonably strong.

5.2. The Probability of Employment

Table 3 reports the estimates of the probit regression on whether native person i in region r at time t is employed in the reference week. We estimate equation (15) separately for native women and native men. These probits are used to compute the inverse Mills ratio included in the individual-level wage regressions discussed below.

The table reports the estimated coefficients on the variables that adjust for differences in reservation wages, such as marital status, presence of young children, and home ownership. We find that marriage lowers the probability of employment for women (by 2 percentage point), but increases it for men (by 10 percentage points). The presence of young children in the household also predicts employment, and the sign of the correlation again differs between men and women. In particular, the presence of young children lowers the probability of employment by 11 percentage points for women, but increases it by 6 percentage points for men. Finally, the probit regressions reveal that household wealth (as proxied by the homeownership variable) has a negative effect on the employment probability for both men and women. Persons who own their home free of debt have a 3 percentage point lower probability of working.[27]

[27] Appendix Table D extends the employment analysis to examine if the immigrant supply shock affects the distribution of full-time employment in the sample of native workers. The probit models reveal a negative relationship between the immigrant share and the probability of full-time employment for native women, but an insignificant impact for native men.

5.3. The Wage of Native Workers

We used the probit regressions to calculate the inverse Mills ratio for each person, and then estimated (separately by gender) the individual-level earnings regressions in equation (14*a*).[28] This exercise produces the selectivity-corrected mean wage of the population in cell (r, t). The cell means then become the dependent variable in equation (14*b*) that examines how immigration affects the wage distribution. Table 4 reports the (OLS and IV) estimated impact of the immigrant supply shock on the adjusted log wage of native women (Panel A) and men (Panel B) at the regional level between 1982 and 2016.

The cell-level regressions are weighted by cell size (i.e., the sum of the individual weights in the cell), and we cluster the standard errors at the region level to account for the possibility of within-group error correlation. Because the number of regions may be too small to estimate the correct cluster-robust standard errors, we implement the wild cluster bootstrap method (Cameron, Gelbach, and Miller, 2008, p. 427) using 1,000 replications and report the corresponding p-values.[29]

Consider initially the results for native women. Column 1 of Table 4 presents the simplest specification, where the (adjusted) mean wage in the cell is calculated from an individual-level regression that does not correct for selection bias and the mean wage in the cell is related only to the immigrant share (plus region and year fixed effects). The OLS coefficient of the immigrant share is insignificant and numerically small, reproducing the descriptive evidence in Figure 2 (which did not adjust for individual differences in education and age).

Column 2 adjusts for selection (i.e., the dependent variable is the region-time fixed effect from an individual-level wage equation that includes the inverse Mills ratio as a regressor). The estimated coefficient of the inverse Mills ratio is strongly positive, suggesting that female workers are positively selected from the female population. The mean value of the inverse Mills ratio for women is 0.47, so that the self-selection of female workers increases the mean of the observed wage distribution by about 8.5 percent

[28] The individual-level regressions in the female (male) sample have 71,337 (103,706) observations.

[29] Cameron, Gelbach and Miller (2008) show that this resampling method provides the most accurate cluster-robust inference in the case of a small number of clusters. Dustmann, Schönberg and Stuhler (2017) and Edo (2020) use this bootstrapping technique in their analysis of the wage impact of immigration.

relative to the population mean (or the product of the coefficient of the inverse Mills ratio and its mean). The OLS coefficient of the immigrant share becomes significantly negative, with a value of -0.44 (0.07). The change in the impact of immigration between columns 1 and 2 is predicted by our theoretical framework if the women who exit the labor force in the post-migration period have relatively low wages.

Columns 5 and 6 present the analogous IV regressions when the immigrant share is instrumented using the shift-share prediction. The OLS and IV coefficients for the simplest model are quite similar. The IV coefficient of the immigrant share in column 5 is insignificant and close to zero, and the coefficient becomes negative and significant (with a value of -0.42, and a standard error of 0.10) when the regression adjusts for selection.

The remaining columns of Table 4 expand the basic model. Columns 3 and 7 do not adjust for selection but add the variable measuring the (log) size of the native labor force. As noted earlier, although the presence of this variable in the equation is implied by the simplest labor demand framework, it has typically been excluded from the regressions estimated in the immigration literature. We limit the discussion to the IV results because of the classic endogeneity issues introduced by this variable.

The log N_{rt} variable has a negative and significant impact on female wages.[30] Note that the wage impact of immigration, as measured by the coefficient of the immigrant share variable, also becomes negative and significant (compared to the simplest model in column 5). The fact that holding constant the size of the native workforce results in a more negative immigration wage effect suggests the existence of a crowd-out effect.

Finally, columns 4 and 8 of Table 4 report the estimates from the full regression specification that controls for both the size of the native labor force and for sample selection. The estimated wage elasticity increases to -0.91 (0.27). In other words, an immigration-induced 10 percent increase in the size of the labor force is predicted to lower the wage of native women by about 9 percent.

[30] The coefficient of the native labor supply variable should equal the coefficient of the immigrant share only if the two groups are perfect substitutes. The sizable difference between the two estimated coefficients may also reflect the fact that our instrument for the native labor supply variable does not fully resolve the endogeneity problems created when higher wages induce more natives to work.

Panel B reports the regressions using the sample of native men. The individual-level estimates from equation (14b) suggest weaker selection for men. The estimated coefficient of the inverse Mills ratio for men is essentially zero. The coefficient of the native labor supply variable is positive, but close to zero and insignificant. As we noted earlier, the instrument for the supply variable $\log N_{rt}$ (based on shift-share projections of the female native population and the presence of small children in the household) may not fully resolve the endogeneity of male labor supply. Note, however, that the coefficient of the immigrant share variable for men is negative, significant, and lies between -0.5 and -0.8, regardless of specification.[31]

In sum, the simplest IV model in column 5 linking immigration and wages suggest a zero correlation between the two variables for women and a negative correlation for men. However, the correction of the biases introduced by the crowd-out effect and the self-selection of workers results in a wage elasticity that has roughly the same value for the two groups. In fact, the difference between the -0.7 elasticity for men and the -0.9 elasticity for women reported in column 8 is not statistically significant (the *t*-statistic is 0.66).[32]

6. Sensitivity to Alternative Selection Corrections

Our baseline model uses a specific combination of family variables and household wealth to calculate the inverse Mills ratio and the selectivity-corrected mean wage. The analysis of female wages includes the family variables and household wealth in the probit regression, and excludes all these variables from the wage regression. The analysis of male wages includes the family variables in both the probit and wage regressions, but includes

---

[31] Although early area studies often found negligible wage effects from immigration (Blau and Mackie, 2016; Edo, 2019), our estimates resemble those reported in more recent studies; see Edo (2020), Ortega and Verdugo (2022), Jaeger, Ruist, and Stuhler (2018), and Monras (2020).

[32] Appendices E, F, and G examine the sensitivity of the evidence reported in Table 4. Appendix E estimates the wage impact when we use the instrument for the immigrant share as the main regressor in a reduced-form equation, when we control for changes in regional industry shares (Bartik, 1991), and when we use alternative specifications of the IV framework. Appendix F shows that the robustness when we use alternative sample periods, measures of the supply shock, regression specifications, samples of native workers, and geographic definitions of the labor market. Finally, Appendix G estimates the impact at the region-skill-year level rather than at the region-year level, and also estimates the wage impact separately for different education groups (showing that the negative wage impact of immigration is particularly affecting low-skill workers).

household wealth only in the probit regression.

Table 5 documents the robustness of the results by estimating alternative selection specifications. Specification 1 reproduces the baseline estimates. Specification 2 includes the family variables in the female individual-level wage regression, but excludes them from the male individual-level wage regression. Specification 3 uses only the family variables and specification 4 uses only the home ownership indicator to generate exogenous variation in the inverse Mills ratio. Finally, the last specification assumes that the error term in the selection equation is uniformly distributed (Olsen, 1980), thereby replacing the inverse Mills ratio with the predicted employment probability calculated from a linear probability model. This specification helps show that the impact of our sample selection correction is not driven by the nonlinearity of the probit model.

While the OLS and IV uncorrected estimates reported in columns 1 and 4 of Panel A are virtually zero for women, adjusting for selection in columns 2 and 5 *always* results in a negative and significant elasticity—regardless of the specification. Columns 3 and 6 add the size of the native labor force to the regression and again show that the estimated wage elasticity for women is robust to the modeling assumptions used (e.g., the IV elasticities range between -0.9 and -1.0 across the five specifications). Panel B shows that the estimated wage elasticities for men are stable across specifications and columns.

We can also bypass the Heckman-type selection correction altogether and show that the estimate of the *uncorrected* wage impact turns negative when we use subsamples of the female workforce where the employment probability is very high. By restricting the analysis to subsamples of women who have high levels of labor force attachment, the estimated wage elasticities are less likely to be contaminated by selection bias. This sampling strategy resembles the "identification-at-infinity" method and does not require the estimation of a first-stage probit that specifies exogenous instruments (Chamberlain, 1986; Heckman, 1990; Mulligan and Rubinstein, 2008).[33]

Table 6 reports the estimated coefficients of the immigrant share variable (using the reduced-form specification of the regression model) for the entire sample of native women

[33] It is important to emphasize that although the identification-at-infinity method corrects for selection bias, it produces results for a very specific group of workers that may not be representative of the entire population (Machado, 2017; Blau, Kahn, Boboshko and Comey, 2021).

and for four alternative subsamples. Specifications 1-2 reports both the uncorrected and selectivity-corrected estimates from Table 4. The relatively low employment rate of native women suggests that selection bias might contaminate the estimated wage impact of immigration.

Specifications 3-5 of Table 6 progressively restrict the sample to women who are likely to have higher levels of labor market attachment, including young women, or young women who do not have young children in the household and who are not married. The employment rate for these subsamples of native women lies between 68 percent and 71 percent (much closer to the employment rate of men of 72.8 percent). The wage elasticities estimated in these subsamples are negative and significant, between -0.3 and -0.5 in column 2. In fact, the size of the elasticity in these subsamples is identical to the selectivity-corrected estimates obtained in the entire sample of native women (equal to -0.4).

Finally, specification 6 estimates the wage impact of immigration by adopting the same strategy as Mulligan and Rubinstein (2008) and Blau, Kahn, Boboshko and Comey (2021). In particular, we first estimate a probit model in each region-time cell since the determinants of employment may vary across regions over time. The dependent variable indicates if the person is employed full-time, and the regressors are the vectors of age and education fixed effects. We then define the identification-at-infinity sample by selecting native women whose predicted employment probability is above the 99th percentile.

By design, the employment rate in this subsample is close to 90 percent. Our analysis uses only 1,421 observations to compute the age- and education-adjusted mean regional wages, approximatively 300 observations in each cross section. Nevertheless, row 6 shows that the identification-at-infinity approach yields a (significant) wage elasticity of about -0.7. In sum, our tests of the sensitivity of the assumptions underlying the selection correction confirm that selection bias contaminates the observed wage impact of immigration for women.[34]

[34] Appendix H tests for the appropriateness of the exclusion restrictions in the typical two-step selection correction used in the baseline model. The exclusion restriction requires that the selection “instruments” (i.e., those variables included in the first-stage probit, but excluded from the second-stage wage regression) are independent of the error term in equation (14*a*). Huber and Mellace (2014) propose a statistical procedure to jointly test the validity of the exclusion restriction (assumption 1) and the assumption that the error in the selection equation is additively separable (assumption 2). The test statistics have large *p*-values, indicating

7. Other Approaches to Identifying the Wage Elasticity

As implied by equation (11), our baseline empirical strategy identifies the wage elasticity η by including the log number of natives as a regressor in the labor demand function. An alternative way of calculating η is to separately estimate the crowd-out effect γ using the employment equation in (12) and then $\eta(1+\gamma)$ using the reduced-form wage equation in (13). This approach avoids instrumenting both the immigrant share and the log size of the native labor force. Table 4 shows that $-\eta\widehat{(1+\gamma)} = -0.44$ for the sample of native women (after accounting for selection bias).

We can estimate γ by regressing the regional employment of native women on the regional immigrant workforce (after dividing both numbers by the working-age population at time t over the 1982-2016 period). Estimating this model and including region and time fixed effects yields an IV estimate of γ equal to -0.51 (with a clustered standard error of 0.06). This estimate implies that the number of female native workers decreases by 5 for every 10 new working immigrants. If we run the same regression in the sample of native men, the estimated γ is -0.004 (with a clustered standard error of 0.06), implying no crowd-out effect.

Combining the separate estimates of γ and $\eta(1+\gamma)$ for the sample of native women implies that η equals -0.90. This numerical value is identical to the one reported in column 8 of Table 4, the specification that also controls for the native labor supply response. The similar results obtained from both methods document the robustness of our baseline empirical strategy.

We can also approximate the extent of selection bias on the wage impact of immigration by using the simulation exercise proposed in Card (2001). This exercise bypasses the estimation of *any* selectivity-corrected wage regressions and instead yields a lower bound for the role of selection bias (if the errors in the market and reservation wage equations are jointly normally distributed).

---

that the tests do not provide evidence against assumptions 1 and 2. We thus fail to reject the validity of the two-step empirical strategy.

Equation (5) of the selection model presented in Section 3 implies that the selection bias on wage growth observed after a supply shock equals $\sigma_\epsilon\, \rho_{\epsilon v}\, [\lambda(\pi_1) - \lambda(\pi_0)]$. To get a numerical estimate of the bias, we need to infer the values of σ_ϵ and $\rho_{\epsilon v}$, and we also need to estimate how much immigration affected π, the female employment rate. Although the standard deviation of the population distribution of earnings σ_ϵ is not observed for women, we assume that the dispersion in earnings is the same for women and men (where there is no selection), implying that $\sigma_\epsilon = 0.43$.

Positive selection into employment in a standard Roy model framework requires $\sigma_\epsilon > \sigma_h$. Let $\sigma_h = \tau\sigma_\epsilon$, where $\tau < 1$. The model of the participation decision in Section 3 then implies that:

$$\rho_{\epsilon v} = \frac{Cov(\epsilon, v)}{\sigma_\epsilon \sigma_v} = \frac{Cov(\epsilon, \epsilon - h)}{\sigma_\epsilon \sigma_v} = \frac{1 - \tau\rho_{\epsilon h}}{\sqrt{1 + \tau^2 - 2\tau\rho_{\epsilon h}}}. \tag{21}$$

For any given value of $0 < \tau < 1$, examination of equation (21) reveals that the correlation $\rho_{\epsilon v}$ has a minimum value in the [0,1] interval. Suppose that $\tau \leq .5$, so that the reservation wage distribution has at most half the dispersion of the market wage distribution. Equation (21) then implies that a lower bound for $\rho_{\epsilon v}$ equals 0.87.

Finally, the bias calculation requires an estimate of the difference $[\lambda(\pi_1) - \lambda(\pi_0)]$. Card (2001) notes that the inverse Mills ratio λ is approximately a linear function of the participation rate π over a wide range of the data. In the French context, the employment rate varies from 0.4 to 0.7 over the 1968-2016 period (see Table 1). A regression of the implied Mills ratio on the employment rate in this range indicates an almost perfectly linear relationship (the regression line is $\lambda = 1.58 - 1.55\pi$, $R^2 = 0.999$). We can then write $[\lambda(\pi_1) - \lambda(\pi_0)] = 1.55(\pi_0 - \pi_1)$.

To measure $\pi_0 - \pi_1$, we calculate the employment response to immigration among native women by using the regional data first illustrated in Figure 2, and estimating an IV regression relating the employment rate between 1982 and 2016 to the immigrant share (controlling for region and time fixed effects). Appendix Table A1 shows that the estimated effect is -0.97 (with a clustered standard error of 0.10), so that a one percentage point increase in the immigrant share reduces the employment rate of native women by 1 percentage point.[35]

Therefore, a lower bound for the selection bias in the observed log wage growth of native women produced by a one percentage point increase in the immigrant share equals $0.43 \times 0.87 \times [1.55 \times 0.97] = 0.56$. In other words, estimates of the wage effect for native women that do *not* correct for selection bias will have a positive bias of about +0.6. The comparison of the various estimates of the wage elasticity reported in Table 4 suggests that selection bias can indeed explain much of the smaller effect estimated in models that ignore the selection problem.

8. Conclusion

An important part of the literature in the economics of immigration examines the impact that immigrants have on the labor market opportunities of native workers in the receiving countries. This research is guided by an intuitive prediction of economic theory: An immigration-induced increase in the size of the labor force should reduce the wage of comparable workers, at least in the short run. Despite the intuitive appeal of this insight, the evidence is mixed, and there is still disagreement on even the direction of the wage impact of immigration after four decades of research on the subject.

Part of the difficulty in measuring the wage impact arises because native workers may respond to the supply shock by moving to labor markets that were not directly affected by immigration. This diffusion of the immigrant supply shock across markets attenuates the wage impact in the targeted market.

This paper explores a hypothesis that provides a deeper understanding into how the diffusion might bias estimates of the wage impact of immigration. The wage change observed in a market targeted by immigrants depends not only on the number of natives who respond, but also on which native workers make the move. A non-random native response changes the composition of the sample of native workers, and this compositional shift artificially changes the average native wage in the affected markets.

[35] Table A1 also shows that estimating the model in the sample of native men yields an insignificant impact of immigration on employment rates. The coefficient is -0.06 (with a clustered standard error of 0.12).

We document the empirical relevance of this selection bias by examining how immigration differentially affected the employment and wages of men and women in France. Beginning with a policy shift in 1976, which gave foreign workers the right to family reunification and made it easier for wives to join their husbands, France experienced a rapid "feminization" of its immigrant workforce. The raw data reveal a gender asymmetry in how immigration correlates with wages and employment. The correlation between immigration and wages (across cities and over time) is negative for native men, but essentially zero for native women. At the same time, the correlation between immigration and employment rates is negative for native women, but essentially zero for native men.

Our theoretical framework combines a labor demand framework with the econometric model of selection to illustrate how the self-selection of the native workforce, and the native response to the immigrant supply shock, contaminates estimates of the key parameters of the labor demand function. Our empirical analysis shows that the orthogonality between immigration and wages for native women is partly an artifact of selection bias. The native women who exited (or did not enter) the labor market tended to be low-wage women, mechanically increasing the average wage in those localities targeted by immigrants and making it seem as if immigration had no impact on the female wage. After adjusting for selection, the wage elasticity for native women is also negative and roughly the same size as that found for native men (where labor supply was inelastic).

The selection bias identified and explored in this paper likely contaminates many existing estimates of the wage impact of immigration. Immigrant supply shocks are likely to have an (immediate) effect on the labor market of receiving countries. Some native workers will respond to these changes in economic opportunities along various margins. The native response is unlikely to be random, altering the composition of the native workforce after the supply shock. A valid assessment of the economic consequences of immigration inevitably requires a thorough examination of the direction and magnitude of the resulting selection bias.

References

Acemoglu, D., Autor, D., & Lyle, D. 2004. Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury. Journal of Political Economy 112(3), 497-555.

Altonji, J. G., & Card, D. 1991. The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives. In Immigration, Trade, and the Labor Market, edited by J. Abowd and R. Freeman. University of Chicago Press, 201-234.

Amior, M. 2020. Immigration, Local Crowd-Out and Undercoverage Bias, Working Paper, Hebrew University of Jerusalem.

Angrist, J.A., & Kugler, A.D. 2003. Protective or Counter-Productive? Labor Market Institutions and the Effect of Immigration on EU Natives. Economic Journal 113 (488), F302-F331.

Aydemir, A., & Borjas, G.J. 2011. Attenuation Bias in Measuring the Wage Impact of Immigration. Journal of Labor Economics, 29(1), 69-112.

Barrett, G.F., & Hamermesh, D.S. 2019. Labor Supply Elasticities: Overcoming Nonclassical Measurement Error Using More Accurate Hours Data. Journal of Human Resources, 54(1), 255-265.

Bartik, Timothy J. 1991. Who Benefits from State and Local Economic Development Policies? WE Upjohn Institute for Employment Research, Kalamazoo (United States).

Blau, F.D., & Kahn, L.M. 2017. The Gender Wage Gap: Extent, Trends, and Explanations. Journal of Economic Literature, 55(3), 789-865.

Blau, F. D., Kahn, L. M., Boboshko, N., & Comey, M. L. 2021. The impact of selection into the labor force on the gender wage gap. Journal of Labor Economics, forthcoming.

Blau, F.D., & Mackie, C., eds. 2016. The Economic and Fiscal Consequences of Immigration. Washington, DC: National Academies Press.

Borjas, G.J. 2003. The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. Quarterly Journal of Economics 118 (4), 1335-1374.

Borjas, G.J. 2006. Native Internal Migration and the Labor Market Impact of Immigration. Journal of Human Resources 41(2), 221-258.

Borjas, G.J., & Monras, J. 2017. The Labour Market Consequences of Refugee Supply Shocks. Economic Policy, 32(91), 361-413.

Borjas, G. J., & Edo, A. (2021). Gender, selection into employment, and the wage impact of immigration (No. w28682). National Bureau of Economic Research.

Bratsberg, B., & Raaum, O. 2012. Immigration and Wages: Evidence from Construction. Economic Journal, 122(565), 1177-1205.

Bratsberg, B., Raaum, O., Røed, M., & Schøne, P. 2014. Immigration Wage Effects by Origin. The Scandinavian Journal of Economics, 116(2), 356-393.

Cameron, A.C., Gelbach, J., & Miller, D. 2008. Bootstrap-Based Improvements for Inference with Clustered Errors. Review of Economics and Statistics 90(3), 414-427.

Card, D. 2001. Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. Journal of Labor Economics, 19(1), 22-64.

Card, D., & Peri, G. 2016. Immigration Economics by George J. Borjas: A Review Essay. Journal of Economic Literature, 54(4), 1333-49.

Chamberlain, G. 1986. Asymptotic efficiency in semi-parametric models with censoring. Journal of Econometrics 32(2), 189-218.

Choi, H.J., Joesch, J.M., & Lundberg, S. 2008. Sons, Daughters, Wives, and the Labour Market Outcomes of West German Men. Labour Economics 15(5), 795-811.

Cortés, P. 2023. Immigration, household production, and native women's labor market outcomes: A survey of a global phenomenon. European Economic Review, 104545.

Dustmann, C., Fabbri, F., & Preston, I. 2005. The impact of immigration on the British labour market. The Economic Journal, 115(507), F324-F341.

Dustmann, C., Frattini, T., & Preston, I.P. 2013. The Effect of Immigration along the Distribution of Wages. Review of Economic Studies, 80(1), 145-173.

Dustmann, C., Schönberg, U., & Stuhler, J. 2017. Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment. Quarterly Journal of Economics, 132(1), 435-483.

Edo, A. 2019. The Impact of Immigration on the Labor Market. Journal of Economic Surveys, 33(3), 922-948.

Edo, A. 2020. The Impact of Immigration on Wage Dynamics: Evidence from the Algerian Independence War. Journal of the European Economic Association 18(6), 3210-3260.

Edo, A., Glesing, Y., Öztunc, J., & Poutvaara, P. 2019. Immigration and Electoral Support for the Far-Left and the Far-Right. European Economic Review 115, 99-143.

Edo, A., & Özgüzel, C. 2023. The impact of immigration on the employment dynamics of European regions. Labour Economics, 85, 102433.

Edo, A., & Toubal, F. 2017. Immigration and the Gender Wage Gap. European Economic Review, 92, 196-214.

Evers, M., De Mooij, R., & Van Vuuren, D. 2008. The Wage Elasticity of Labour Supply: A Synthesis of Empirical Estimates. De Economist, 156(1), 25-43.

Foged, M. & Peri, G. 2016. Immigrants' Effect on Native Workers: New Analysis on Longitudinal Data. American Economic Journal: Applied Economics 8(2), 1-34.

Gardner, J. 2020. Immigration Displaces Women. Working Paper, University of Mississippi.

Glitz, A. 2012. The Labor Market Impact of Immigration: A Quasi-Experiment Exploiting Immigrant Location Rules in Germany, Journal of Labor Economics 30(1), 175–213.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. 2020. Bartik instruments: What, when, why, and how. American Economic Review, 110(8), 2586-2624.

Gronau, R. 1974. Wage Comparisons: A Selectivity Bias. Journal of Political Economy 82(6), 1119-43.

Gurgand, M., & Margolis, D.N. 2008. Does Work Pay in France? Monetary Incentives, Hours Constraints, and the Guaranteed Minimum Income. Journal of Public Economics, 92(7), 1669-1697.

Heckman, J. 1979. Sample Selection Bias as a Specification Error, Econometrica 47(1), 153–161.

Heckman, J. 1990. Varieties of selection bias. American Economic Review, 80(2), 313-318.

Huber, M. & Mellace, G. 2014. Testing Exclusion Restrictions and Additive Separability in Sample Selection Models, Empirical Economics 47, 75-92.

Jaeger, D. A., Ruist, J., & Stuhler, J. 2018. Shift-Share Instruments and the Impact of Immigration (No. w24285). National Bureau of Economic Research.

Kleibergen, F., & Paap, R. 2006. Generalized Reduced Rank Tests Using the Singular Value Decomposition. Journal of Econometrics 133(1), 97-126.

Laroque, G., & Salanié, B. 2002. Labour Market Institutions and Employment in France. Journal of Applied Econometrics, 17(1), 25-48.

Llull, J. 2018. Immigration, Wages, and Education: A Labour Market Equilibrium Structural Model. Review of Economic Studies, 85(3), 1852-1896

Llull, J. 2021. Immigration and Gender Differences in the Labor Market. Journal of Human Capital, 15(1), 174-203.

Lundberg, S., & Rose, E. 2000. Parenthood and the Earnings of Married Men and Women. Labour Economics 7(6), 689-710.

Maasoumi, E., & Wang, L. 2019. The gender gap between earnings distributions. Journal of Political Economy, 127(5), 2438-2504.

Machado, C. 2017. Unobserved selection heterogeneity and the gender wage gap. Journal of Applied Econometrics, 32(7), 1348-1366.

McDonald, P. 2020. The Male Marriage Premium: Selection, Productivity, or Employer Preferences? Journal of Marriage and Family, forthcoming.

Monras, J. 2020. "Immigration and Wage Dynamics: Evidence from the Mexican Peso Crisis," Journal of Political Economy 128 (August), 3017-3089.

Monras, J. 2021. Local Adjustment to Immigrant-driven Labor Supply Shocks, Journal of Human Capital 15(1), 204-35.

Mulligan, C.B., & Rubinstein, Y. 2008. Selection, Investment, and Women's Relative Wages over Time, Quarterly Journal of Economics 123 (3), 1061-1110.

Muthén, Bengt. 1990. Moments of the Censored and Truncated Bivariate Normal Distribution, British Journal of Mathematical and Statistical Psychology 43(1), 131-143.

Olsen, R.J. 1980. A least squares correction for selectivity bias. Econometrica 48(7), 1815-1820.

Ortega, J., & Verdugo, G. 2022. Who stays and who leaves? Immigration and the selection of natives across locations. Journal of Economic Geography, 22(2), 221-260.

Pencavel, J. 1986. Labor Supply of Men: A Survey. In Handbook of Labor Economics, Volume 1, edited by O. Ashenfelter and R. Layard. Elsevier Science Publishers, 3-102.

Piketty, T. 1998. L'impact des Incitations Financières au Travail sur les Comportements Individuels: Une Estimation pour le cas Français. Économie & prévision, 132(1), 1-35.

Rosenbaum, S. 1961. Moments of a Truncated Bivariate Normal Distribution. Journal of the Royal Statistical Society, Series B (Methodological) 23(2), 405-408.

Sanderson E., & Windmeijer, F. 2016. A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables," Journal of Econometrics 190(2), 212-221.

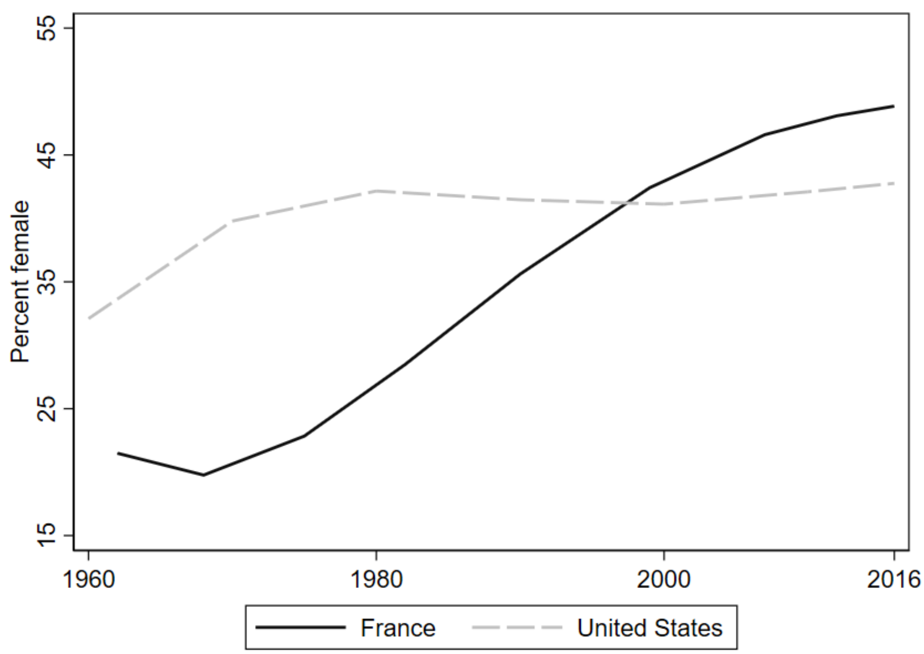
Stock, J., Yogo, M. & Wright, J. 2002. A Survey of Weak Instruments and Weak Identification in a Generalized Method of Moments," Journal of Business and Economic Statistics 20(4), 518-529.

Topel, R. H. 1997. Factor proportions and relative wages: the supply-side determinants of wage inequality. Journal of Economic perspectives, 11(2), 55-74.

Figure 1. Immigration and gender

A. Trends in the immigrant share in the French labor force

B. The feminization of the immigrant labor force, France v. USA



Source: INSEE, French censuses; IPUMS, USA decennial censuses and American Community Surveys.

Figure 2. Immigration, wages, and employment of native men and women

Figure 2A

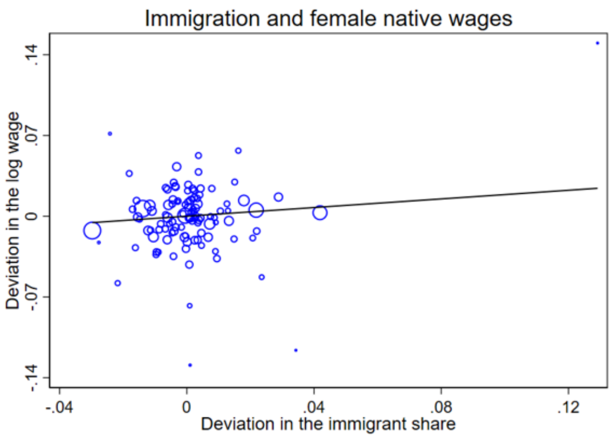


Figure 2B

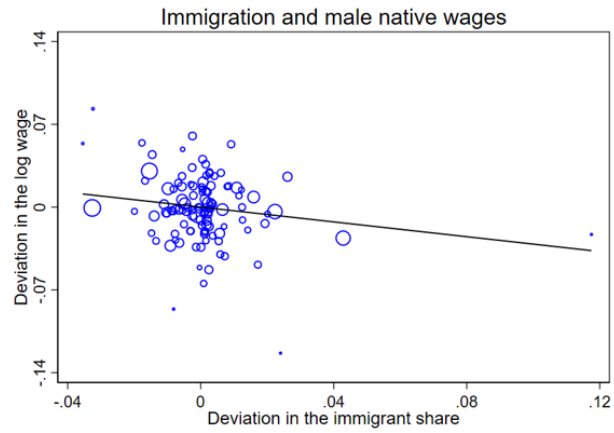


Figure 2C

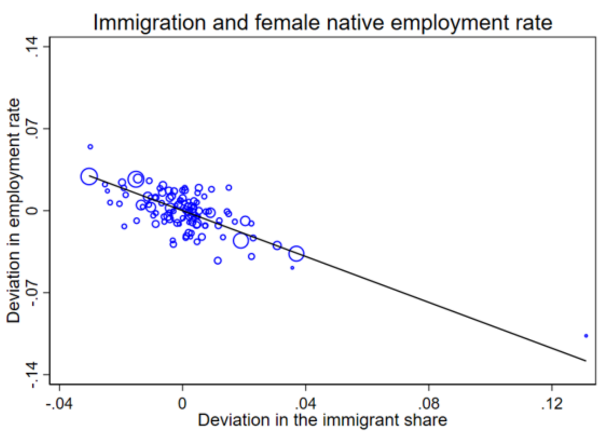
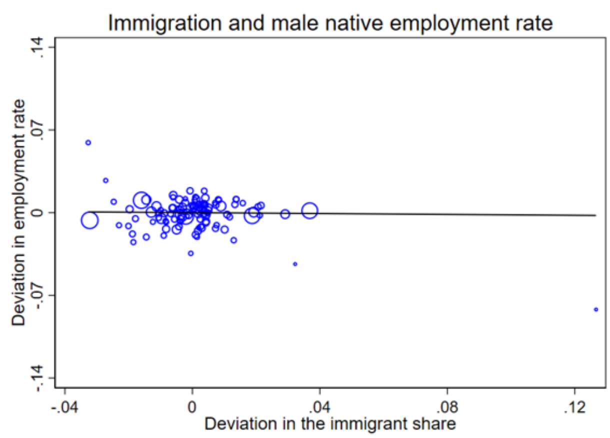


Figure 2D



Notes: The unit of observation in the scatter diagrams is a region-year cell over the 1982-2016 period. Figures 2A and 2B (Figures 2C and 2D) correlate the deviation in the log monthly wage (employment rate) of native women and men, respectively, to the deviation in the immigrant share after removing any year-specific effects that are common to all regions in a given census year. The deviations in the log wage, employment rate, or immigrant share are residuals from regressions of these variables on region fixed effects and census year fixed effects. The regression line in the figures weights the data by the number of observations used to compute the dependent variable. The size of the circles is proportional to the value of the weight.

Figure 3. Selection bias and the wage impact of immigration

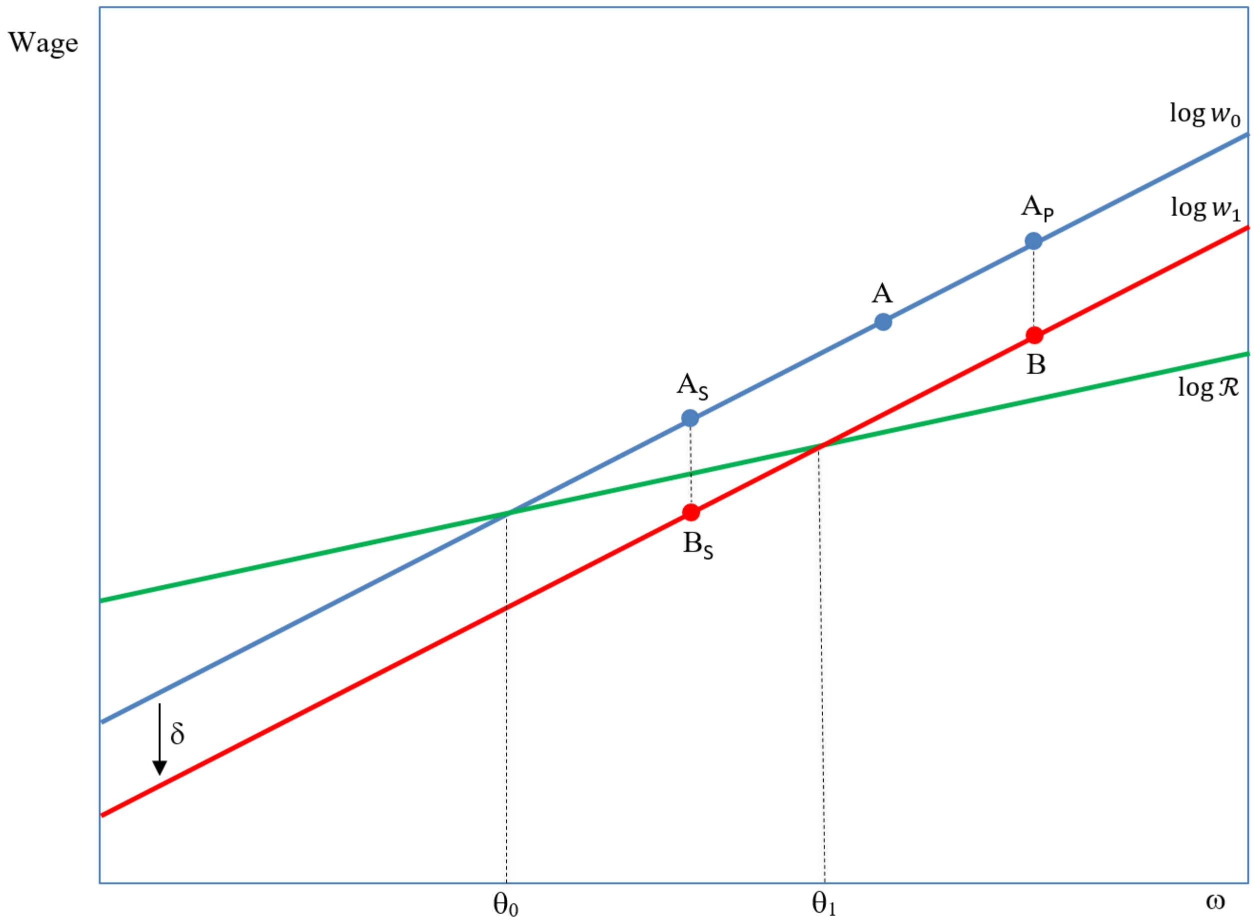


Table 1: Descriptive statistics

	1962	1968	1975	1982	1990	1999	2007	2016
	A. French census data							
Employment rate of female natives	37.06	40.08	47.55	51.51	56.40	62.19	67.76	70.06
Employment rate of male natives	89.30	87.36	86.62	81.05	77.34	75.47	75.45	73.64
Immigrant share	7.71	9.75	10.33	9.21	8.86	8.76	9.71	11.31
Immigrant share, women	5.07	5.68	6.25	6.27	6.93	7.79	9.18	10.96
With a baccalaureate degree	2.78	2.94	3.05	3.25	4.58	5.82	7.38	9.19
With less than a baccalaureate degree	5.37	6.17	7.19	7.37	8.08	9.33	11.38	14.13
Immigrant share, men	8.99	11.83	12.81	11.32	10.48	9.64	10.22	11.66
With a baccalaureate degree	4.85	4.55	4.66	4.83	6.50	7.80	8.86	9.95
With less than a baccalaureate degree	9.47	12.96	14.63	13.17	11.99	10.69	11.4	13.78
Share of natives with a baccalaureate degree	11.14	14.92	21.37	25.35	31.01	40.83	51.65	60.93
Women	10.85	14.58	19.97	23.75	28.72	37.14	47.39	56.34
Men	11.71	15.52	23.52	27.45	33.64	44.81	55.99	65.40
	B. French labor force survey data							
Average wage of female natives	-	-	-	1626.6	1684.2	1758.3	1842.0	1890.0
Average wage of male natives	-	-	-	2077.8	2091.2	2100.58	2180.6	2236.8
Employment rate of female natives	-	-	-	55.18	56.60	61.73	63.57	65.83
Employment rate of male natives	-	-	-	83.41	78.69	76.28	71.55	70.42
Observations	-	-	-	32,446	78,531	83,311	59,414	75,446

Notes: The table uses data drawn from the French censuses (Panel A) and the French Labor Force Surveys (Panel B). The immigrant shares are computed using the sample of persons in the labor force and are defined as $\log(1 + M/N)$, where M and N give the number of foreign-born and native labor force participants, respectively. The share of natives with a baccalaureate degree is computed using the sample of persons in the labor force.

Table 2: Instrumental variables, first-stage regressions

	Sample of native women		Sample of native men	
	(1)	(2)	(3)	(4)
	A. Single endogenous variable model			
Dependent variable: Immigrant share				
Predicted immigrant share in population	1.77***	-	1.71***	-
	(0.31)		(0.37)	
Kleibergen-Paap F-test of excluded instrument	32.06	-	21.00	-
	B. Two endogenous variables model			
Dependent variable: Immigrant share				
Log predicted immigrant population	-	0.12***	-	0.11***
		(0.02)		(0.03)
Log predicted female native labor force	-	-0.14***	-	-0.13***
		(0.04)		(0.04)
SW multivariate F-test of excluded instruments	-	12.91	-	14.09
Dependent variable: Log of native labor force				
Log predicted immigrant population	-	-0.09	-	-0.09
		(0.08)		(0.08)
Log predicted female native labor force	-	0.58***	-	0.58***
		(0.09)		(0.09)
SW multivariate F-test of excluded instruments	-	15.15	-	16.53

Notes: Standard errors reported in parentheses are heteroscedasticity robust and clustered by region. The table reports the first-stage IV regressions in the estimation sample. In Panel A, the dependent variable is the immigrant share in the labor force. In Panel B, the dependent variables are the immigrant share and the log number of natives in the labor force. As instruments, we use the predicted immigrant share in the population based on the geographic settlement of immigrants and natives in the 1968 census and the predicted female native labor force based on the geographic settlement of natives in the 1968 census and the relative number of women with young children in subsequent years. As tests for weak instruments, Panel A reports the Kleibergen-Paap rk Wald F-test for the excluded instrument, while Panel B reports the Sanderson-Windmeijer (SW) F-tests of excluded instruments for each endogenous regressor. All regressions include region and time fixed effects, and are weighted by cell size (i.e., the sum of individual weights in a cell). ***, **, * denote statistical significance from zero at the 1%, 5%, 10% significance level.

Table 3: Probit regressions on the employment probability of natives

	Native women	Native men
	(1)	(2)
Married	-0.06**	0.42***
	(0.02)	(0.02)
Marginal effect	*-0.02*	*0.10*
Presence of children below 6	-0.37***	0.25***
	(0.02)	(0.02)
Marginal effect	*-0.11*	*0.06*
Home ownership	-0.11***	-0.12***
	(0.02)	(0.02)
Marginal effect	*-0.03*	*-0.03*
Observations	173,432	155,716

Notes: Standard errors reported in parentheses are heteroscedasticity robust and clustered by region. Below the standard errors, we report the marginal effect of each variable computed at the mean value of the sample. The dependent variable is a binary variable equal to one if the individual is employed and zero otherwise. All regressions include age, education, region-time fixed effects, and use the individual weight provided by INSEE. ***, **, * denote statistical significance from zero at the 1%, 5%, 10% significance level.

Table 4: Impact of immigration on native wages

	OLS estimates				IV estimates			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Impact on the wage of native women								
Immigrant share	-0.09	-0.44***	-0.07	-0.37**	-0.08	-0.42***	-0.63**	-0.91***
	(0.07)	(0.07)	(0.13)	(0.14)	(0.10)	(0.10)	(0.28)	(0.27)
Wild cluster bootstrap p-value	*0.19*	*0.01*	*0.51*	*0.03*	*0.47*	*0.06*	*0.10*	*0.00*
Log of native labor force	-	-	0.01	0.04	-	-	-0.23***	-0.20**
			(0.07)	(0.06)			(0.09)	(0.08)
Wild cluster bootstrap p-value			*0.82*	*0.51*			*0.09*	*0.09*
Selectivity-corrected estimates	-	Yes	-	Yes	-	Yes	-	Yes
Inverse Mills ratio	-	0.18***	-	0.18***	-	0.18***	-	0.18***
		(0.02)		(0.02)		(0.02)		(0.02)
Kleibergen-Paap F-test	-	-	-	-	32.06	32.06	-	-
SW multivariate F-test (imm. share)	-	-	-	-	-	-	12.91	12.91
SW multivariate F-test (log nat.)	-	-	-	-	-	-	15.15	15.15
B. Impact on the wage of native men								
Immigrant share	-0.70***	-0.70***	-0.55***	-0.54***	-0.81***	-0.81***	-0.70***	-0.70***
	(0.12)	(0.12)	(0.17)	(0.17)	(0.09)	(0.09)	(0.17)	(0.17)
Wild cluster bootstrap p-value	*0.21*	*0.21*	*0.04*	*0.04*	*0.00*	*0.00*	*0.01*	*0.01*
Log of native labor force	-	-	0.09	0.09	-	-	0.03	0.02
			(0.06)	(0.06)			(0.06)	(0.06)
Wild cluster bootstrap p-value			*0.15*	*0.15*			*0.70*	*0.73*
Selectivity-corrected estimates	-	Yes	-	Yes	-	Yes	-	Yes
Inverse Mills ratio	-	-0.01	-	-0.01	-	-0.01	-	-0.01
		(0.07)		(0.07)		(0.07)		(0.07)
Kleibergen-Paap F-test	-	-	-	-	21.00	21.00	-	-
SW multivariate F-test (imm. share)	-	-	-	-	-	-	14.09	14.09
SW multivariate F-test (log nat.)	-	-	-	-	-	-	16.53	16.53

Notes: Standard errors reported in parentheses are heteroscedasticity robust and clustered by region. The unit of observation is a region-year cell over the 1982-2016 period, and all regressions have 110 observations (22 regions and 5 years). The dependent variable is the age- and education-adjusted wage of native women (Panel A) or men (Panel B). Columns 2, 4, 6 and 8 further adjust wages for sample selection. Columns 5-6 instrument the share of immigrants with the shift-share instrument computed using the 1968 French census; columns 7-8 instrument both the share of immigrants and the log native labor force by using the shift-share instrument and the predicted (log) size of the female native labor force. All regressions include region and time fixed effects, and are weighted by cell size. Wild bootstrap p-values in italics are computed using 1,000 bootstrap replications. ***, **, * denote statistical significance from zero at the 1%, 5%, 10% significance level.

Table 5: Immigration and wages using alternative selection models

	Regressors included: X = age-time, education-time, region f.e.; F = family characteristics; H = home ownership		OLS estimates			IV estimates		
			(1)	(2)	(3)	(4)	(5)	(6)
	Employment regression	Individual-level wage regression	**A. Impact on the wage of native women**					
Baseline specification	(X,F,H)	(X)	-0.09	-0.44***	-0.37**	-0.08	-0.42***	-0.91***
			(0.07)	(0.07)	(0.14)	(0.10)	(0.10)	(0.27)
Wild cluster bootstrap p-value			*0.19*	*0.01*	*0.03*	*0.47*	*0.06*	*0.00*
Specification 2	(X,F,H)	(X,F)	-0.09	-0.38***	-0.32**	-0.08	-0.36***	-0.87***
			(0.07)	(0.07)	(0.14)	(0.10)	(0.10)	(0.27)
Wild cluster bootstrap p-value			*0.19*	*0.02*	*0.04*	*0.47*	*0.07*	*0.00*
Specification 3	(X,F)	(X)	-0.09	-0.40***	-0.34**	-0.08	-0.38***	-0.88***
			(0.07)	(0.07)	(0.14)	(0.10)	(0.10)	(0.27)
Wild cluster bootstrap p-value			*0.19*	*0.02*	*0.04*	*0.47*	*0.07*	*0.00*
Specification 4	(X,H)	(X)	-0.09	-0.54***	-0.45***	-0.08	-0.52***	-0.98***
			(0.07)	(0.08)	(0.14)	(0.10)	(0.10)	(0.27)
Wild cluster bootstrap p-value			*0.19*	*0.01*	*0.01*	*0.47*	*0.04*	*0.00*
Baseline specification using linear probability model	(X,F,H)	(X)	-0.09	-0.49***	-0.42***	-0.08	-0.47***	-0.94***
			(0.07)	(0.08)	(0.14)	(0.10)	(0.10)	(0.27)
Wild cluster bootstrap p-value			*0.19*	*0.01*	*0.02*	*0.47*	*0.04*	*0.00*
	Employment regression	Individual-level wage regression	**B. Impact on the wage of native men**					
Baseline specification	(X,F,H)	(X,F)	-0.70***	-0.70***	-0.54***	-0.81***	-0.81***	-0.70***
			(0.12)	(0.12)	(0.17)	(0.09)	(0.09)	(0.17)
Wild cluster bootstrap p-value			*0.21*	*0.21*	*0.04*	*0.00*	*0.00*	*0.01*
Specification 2	(X,F,H)	(X)	-0.64***	-0.41***	-0.30	-0.76***	-0.53***	-0.81***
			(0.12)	(0.11)	(0.20)	(0.09)	(0.11)	(0.25)
Wild cluster bootstrap p-value			*0.23*	*0.11*	*0.18*	*0.00*	*0.02*	*0.00*
Specification 3	(X,F)	(X)	-0.64***	-0.39***	-0.29	-0.76***	-0.52***	-0.82***
			(0.12)	(0.11)	(0.21)	(0.09)	(0.11)	(0.25)
Wild cluster bootstrap p-value			*0.23*	*0.12*	*0.20*	*0.00*	*0.03*	*0.00*
Specification 4	(X,H)	(X)	-0.64***	-0.68***	-0.52***	-0.76***	-0.79***	-0.63***
			(0.12)	(0.12)	(0.17)	(0.09)	(0.09)	(0.18)
Wild cluster bootstrap p-value			*0.23*	*0.27*	*0.05*	*0.00*	*0.00*	*0.01*
Baseline specification using linear probability model	(X,F,H)	(X,F)	-0.70***	-0.93***	-0.74***	-0.81***	-1.03***	-0.61***
			(0.12)	(0.16)	(0.17)	(0.09)	(0.13)	(0.22)
Wild cluster bootstrap p-value			*0.21*	*0.42*	*0.02*	*0.00*	*0.01*	*0.04*
Selectivity-corrected estimates			-	Yes	Yes	-	Yes	Yes
Add log of native labor force as regressor			-	-	Yes	-	-	Yes

Notes: Standard errors reported in parentheses are heteroscedasticity robust and clustered by region. The unit of observation is a region-year cell over the 1982-2016 period, and all regressions have 110 observations (22 regions and 5 years). The dependent variable is the age- and education-adjusted wage of native women (Panel A) or men (Panel B). Columns 2-3 and 5-6 further adjust wages for sample selection. Each row uses a specific set of variables to generate the inverse Mills ratio and estimate the wage regressions. Columns 4-5 instrument the share of immigrants with the shift-share instrument computed using the 1968 French census; column 6 instruments both the share of immigrants and the log native labor force by using the shift-share instrument and the predicted (log) size of the female native labor force. All regressions include region and time fixed effects, and are weighted by cell size. Wild bootstrap p-values in italics are computed using 1,000 bootstrap replications. ***, **, * denote statistical significance from zero at the 1%, 5%, 10% significance level.

Table 6: Identification-at-infinity using alternative samples of female native workers

	Employment rate to population	# obs. to compute regional wages	OLS (1)	IV (2)
1. Native women	63.5%	71,337	-0.09	-0.08
			(0.07)	(0.10)
Wild cluster bootstrap p-value			*0.19*	*0.47*
Kleibergen-Paap F-test			-	32.06
2. Native women: Selectivity-corrected estimates	63.5%	71,337	-0.44***	-0.42***
			(0.07)	(0.10)
Wild cluster bootstrap p-value			*0.01*	*0.06*
Kleibergen-Paap F-test			-	32.06
3. Native women, aged 21-35	67.7%	27,634	-0.22***	-0.29***
			(0.09)	(0.11)
Wild cluster bootstrap p-value			*0.16*	*0.07*
Kleibergen-Paap F-test			-	32.49
4. Native women without young children, aged 21-35	70.8%	18,731	-0.39***	-0.47***
			(0.11)	(0.13)
Wild cluster bootstrap p-value			*0.06*	*0.06*
Kleibergen-Paap F-test			-	33.91
5. Native single women without young children, aged 21-35	69.9%	13,651	-0.40***	-0.40***
			(0.12)	(0.12)
Wild cluster bootstrap p-value			*0.15*	*0.06*
Kleibergen-Paap F-test			-	41.26
6. Native women, $\boldsymbol{Pr(\widehat{EMP}=1) > 99th\ percentile}$	89.5%	1,421	-0.34	-0.71**
			(0.56)	(0.35)
Wild cluster bootstrap p-value			*0.44*	*0.29*
Kleibergen-Paap F-test			-	48.72

Notes: Standard errors reported in parentheses are heteroscedasticity robust and clustered by region. The unit of observation is a region-year cell over the 1982-2016 period. Specifications 1-5 exploit 110 observations (22 regions and 5 years), and specification 6 uses 28 observations. The dependent variable is the age- and education-adjusted wage of native women in specification 1 and 3-6. Specification 2 further adjust wages for sample selection. Column 2 instruments the share of immigrants with the shift-share instrument computed using the 1968 French census. All regressions include region and time fixed effects, and are weighted by cell size. Wild bootstrap p-values in italics are computed using 1,000 bootstrap replications. ***, **, * denote statistical significance from zero at the 1%, 5%, 10% significance level.