## Occupational Barriers and the Productivity Penalty from Lack of Legal Status<sup>\*</sup>

Francesc Ortega<sup>†</sup> CUNY, Queens College

Amy Hsin<sup>‡</sup> CUNY, Queens College

May 12, 2022

#### Abstract

Wage gaps between documented and undocumented workers reflect employer exploitation, endogenous occupational sorting and productivity losses associated with lack of legal status. Our paper presents a model-based strategy to identify the productivity penalty associated with lack of legal status, which is crucial to estimate the net economic gains from legalization. In the model, heterogeneous workers choose occupations and undocumented workers are subject to employer discrimination and experience productivity loss in occupations characterized by tasks that require legal status. The theoretical analysis provides guidance on how to identify occupational barriers and delivers an easy-to-compute lower bound for the undocumented productivity penalty. Applying this strategy to individuallevel data that imputes undocumented status, we estimate that the productivity penalty associated with lack of legal status in the United States is upward of 12%and affects roughly one third of undocumented workers, which in turn account for over 5% of US employment. Thus, legalization of undocumented workers would not only improve their wages, but also increase GDP by a minimum of 0.96% per year.

JEL Classifications: J3, J7

*Keywords*: Immigration; Undocumented; Legalization; Discrimination

<sup>\*</sup>We thank the William T. Grant Foundation and PSC/CUNY for generous financial support. We also thank Catalina Amuedo-Dorantes, Marianne Bertrand, Janet Calvo, Michael Clemens, Albrecht Glitz, Laura Giuliano, Jennifer Hunt, Fernando Lozano, Lilia Maliar, Joan Monras, Giovanni Peri, Amy Ellen Schwartz, Yochanan Shachmurave, Kevin Shih, Todd Sorensen, Chad Sparber, Suleyman Taspinar, Thom Thurston, Marcos Vera-Hernandez and Wim Vijverberg for helpful comments. We are also grateful to the Editor and two anonymous referees for their suggestions.

<sup>&</sup>lt;sup>†</sup>Francesc Ortega is the Dina A. Perry Professor of Economics, Queens College, CUNY. E-mail: francesc.ortega@qc.cuny.edu. Homepage: (http://qcpages.qc.cuny.edu/~fortega).

<sup>&</sup>lt;sup>‡</sup>Associate Professor of Sociology, Queens College, CUNY. amy.hsin@qc.cuny.edu. Homepage: (https://sites.google.com/view/amyhsin)

## 1 Introduction

Many countries are home to large numbers of unauthorized immigrants.<sup>1</sup> Despite lacking the right to reside or work legally, unauthorized immigrants contribute in significant ways to the economies of the host countries. Roughly 11 million unauthorized immigrants live in the United States, comprising 5% of the labor force and contributing over 3% of GDP (Edwards and Ortega, 2017). Legalization of undocumented workers is widely debated by policy-makers and social scientists. While questions of human rights and ethics are foundational to these debates, so are questions related to the economic effects of legalization for host countries.

A large body of literature shows that the wages and working conditions of undocumented immigrants increase when they gain legal status. In the context of the United States, many studies have supported this claim based on the 1986 IRCA legalization (Rivera-Batiz, 1999; Kossoudji and Cobb-Clark, 2002; Amuedo-Dorantes et al., 2007; Lozano and Sorensen, 2011; Pan, 2012) and, more recently, on the 2012 DACA program providing temporary work permits to undocumented youth (Pope, 2016; Amuedo-Dorantes and Antman, 2017).

However, the previous evidence is insufficient to answer some key concerns in the debate about the economic effects of legalization, such as the aggregate effects on GDP and on government coffers. Answering these questions requires distinguishing how much of the wage increase upon gaining legal status can be attributed to a gain in productivity versus other factors, such as the loss of employers' ability to exploit undocumented workers. While the latter mainly entails income redistribution from employers to formerly undocumented workers, productivity increases generate a net increase in income (and tax revenue) for the host country. Quantifying the undocumented productivity penalty is crucial in structural analyses aimed at estimating the net economic contribution of undocumented workers and simulating the effects of legalization on GDP, the wage structure and government coffers (Edwards and Ortega, 2017; Machado, 2017; Ortega et al., 2019; Peri and Zaiour, 2021)).

Identifying the productivity gains associated with gaining legal status is a challenging task. While hard to quantify with precision, several studies have shown that unauthorized immigrants suffer wage exploitation (Gleeson and Gonzales, 2012; Brown et al., 2013; Bartolucci, 2014). At the same time, there is clear evidence that (implicit) occu-

 $<sup>^1{\</sup>rm The}$  United Nations Development Programme estimated the worldwide unauthorized immigrant population to be over 50 million people in 2009.

pational barriers lead to occupational mismatch and diminish worker productivity (Weeden, 2002; Hsieh et al., 2019). The labor market opportunities of undocumented workers are almost certainly diminished by occupational barriers in similar ways (Abrego, 2011; Amuedo-Dorantes and Antman, 2017). These barriers vary importantly across occupations, reflecting regulatory constraints, such as legal residence requirements associated with licenses, as well as the nature of the specific tasks involved in each occupation. For instance, the need to hold face-to-face interactions with customers or government agencies, or to travel extensively, exposes undocumented workers to apprehension and deportation. Besides reducing the productivity of undocumented workers in these occupations, these entry barriers are likely to distort their occupational choices.

Our paper presents a new strategy to identify the productivity penalty associated with lack of legal status. We lay out a theoretical model where heterogeneous workers choose occupations (as in the Roy model). Some occupations entail tasks that require legal status. As a result, undocumented workers in these occupations suffer a productivity loss that entails lower wages and acts as an entry barrier into those occupations and distorts their occupational choices. In addition, employers may exploit undocumented workers in all occupations, paying them wages below productivity. The theoretical analysis suggests an empirical strategy to identify which occupations have entry barriers for undocumented workers, clarifies the factors that determine the productivity and wage gaps between documented and undocumented workers, and shows how to estimate the undocumented productivity penalty. Additionally, the model also illustrates the labor market effects of legalization in terms of occupational switching, wage growth and net economic gains. An important lesson of our analysis is that exact identification of the undocumented productivity penalty is infeasible due to endogenous occupational sorting in terms of unobserved idiosyncratic productivity. However, even in this scenario, we derive a *lower bound* for the productivity loss associated with lack of legal status.

The second part of the paper goes on to implement this strategy using a special extract of the *American Community Survey* (ACS) that also includes a sophisticated imputation to identify likely undocumented individuals. Our empirical analysis has two main findings. First, we identify the occupations with the largest entry barriers to undocumented workers. Many of these occupations require legal status (e.g. teachers and nurses) or entail tasks that involve driving, long-distance travel or face-to-face interaction with the public and government officials (e.g. managers, secretaries or salespersons). These tasks cannot be accomplished in full by undocumented workers, reducing their productivity in these jobs and distorting their career choices. Secondly, we estimate that

the productivity penalty associated with lack of legal status is upwards of 12 percent and affects roughly one third of all undocumented workers. This finding implies that legalization would increase GDP and we quantify this increase to be *at least* 0.96% per year.

Our analysis is not only relevant in the United States. Unauthorized immigration is pervasive in high-income countries that are in geographical proximity to countries with demographic, economic or political pressures (Orrenius and Zavodny, 2016). Several studies have used European data to analyze the economic effects of legalization. For instance, Monras et al. (2017) empirically analyze a large legalization process in Spain. Among other findings, they show that legal status increased the labor market opportunities of immigrants. Along similar lines, Devillanova et al. (2018) study the effect of the *prospect* of legal status on the employment of undocumented immigrants in Italy, finding a positive effect. Inevitably, an important factor in the discussions on whether to provide legal status to undocumented workers in receiving countries is the consequences of such a policy for GDP and the public coffers. As argued above, these effects rely crucially on whether legal status increases the productivity of undocumented workers or simply redistributes income from employers to employees.

The structure of the paper is the following. Section 2 summarizes the relevant literature. Section 3 presents our theoretical analysis. Section 4 presents the data and descriptive statistics. Section 5 estimates the gaps in occupational shares between documented and undocumented workers, Section 6 estimates the undocumented productivity penalty, and Section 7 concludes.

## 2 Related Literature

Our paper draws on the broad literature analyzing wage gaps by race and gender and applies a similar approach to estimating wage gaps by legal status. One of the most relevant studies for our paper is Hsieh et al. (2019), which argues that occupational barriers led to substantial misallocation of talent by race and gender in the United States. The authors use a generalized Roy (1951) model where individuals first choose education and later enter the labor market by choosing occupations. These groups face barriers to human capital accumulation and occupational choice to different degrees. In our paper, occupational barriers also play a central role as a determinant of the occupational choices of the minority group (i.e. undocumented workers), but our model is static and leaves out distortions to human capital accumulation, which will understate the overall long-run effects of undocumented status on individual and aggregate income.

Our paper is also connected to the literature on the labor market outcomes of undocumented workers. This literature demonstrates a large wage differential between documented and undocumented workers with similar skills. Using the Survey of Income and Program Participation, Hall et al. (2010) estimated a 17% wage gap between documented and undocumented male Mexicans. A number of studies have examined the wage effects of the 1986 IRCA amnesty, estimating undocumented wage penalties ranging between 5% and 20% (Rivera-Batiz, 1999; Kossoudji and Cobb-Clark, 2002; Amuedo-Dorantes et al., 2007; Lozano and Sorensen, 2011; Pan, 2012)). Orrenius and Zavodny (2015) provide additional evidence of the wage penalty associated with undocumented status by showing that the introduction of E-Verify, a program that allows employers to verify the legal status of employees, led to a reduction on the wages of undocumented workers.

More recently, Albert (2021) documents differences in wages and job finding rates by legal status. Using data from the *Current Population Survey* (CPS), he imputes legal status and estimates that, conditional on observable characteristics, undocumented immigrants earn 8% less and have a 7 percentage-point higher probability of finding a job than natives. In his model, firms prefer workers with lower bargaining power (because they can extract more surplus) and are able to discriminate between hiring native and immigrant workers, extending (Chassambouli and Peri, 2015). Borjas and Cassidy (2019) also produced estimates of the wage gaps between observationally equivalent immigrants differing in documentation status. They impute documentation status in the 2008-2016 waves of the American Community Survey (ACS) with an approach similar to the one used in our data. Similar to our own findings, they find that (in 2012-2013) the wages of undocumented workers were roughly 6 log points lower than for documented workers with comparable education and demographic information. Importantly, the longer time span in their data allows them to describe the evolution over time in the documentation wage gap, which shows a reduction in the wage gap after the implementation of the DACA program, which provided temporary legal status to undocumented youth who arrived to the United States as children (also known as Dreamers).

In the last few years, several studies have focused on the effects of DACA on the labor market and educational outcomes of Dreamers. Pope (2016) and Amuedo-Dorantes and Antman (2017) use data from the ACS and CPS, respectively. Lacking information on immigrants' legal status, these authors were forced to assume that non-citizens in a given age range are undocumented. Both studies find positive effects of DACA on employment, but disagree on the effects on schooling. Hsin and Ortega (2018) use administrative data that allows for a precise identification of students' legal status. They find that DACA led to a large increase in dropout rates among undocumented college students enrolled at 4year colleges (though not among those attending community college). In a recent study, Kuka et al. (2020) provide evidence that DACA incentivized human capital investments among teenagers. In comparison, our study uses data for the period immediately prior to DACA and focuses on the quantification of productivity loss associated with lack of legal status.

As noted earlier, the positive effect of legalization on the wages of undocumented workers does not necessarily imply an increase in their productivity and, consequently, on overall GDP. It might simply reflect the strengthening in these workers' bargaining power and the enhanced ability to enforce their rights. Undocumented immigrants reside in the country without work authorization and can be deported, which makes them vulnerable to employer exploitation. Through qualitative analysis, Gleeson and Gonzales (2012) find that undocumented workers are commonly subjected to workplace violation of labor laws and are deterred from filing complaints because they fear employers will retaliate by reporting them to immigration authorities. Brown et al. (2013) analyze administrative data from Georgia state and identify which firms employ undocumented workers on the basis of erroneous social security numbers. The results suggest that firms with undocumented workers experience a competitive advantage, which translates into a higher rate of survival. Undocumented workers are also more likely to work in jobs that are physically strenuous and hazardous and receive no compensating differentials for working in unfavorable environments (Hall and Greenman, 2015).

Nonetheless, a number of studies have documented that illegality negatively affects worker productivity in multiple ways. For instance, the threat of deportation and heavily restricted labor market opportunities increases the risk of depression and anxiety among undocumented youth (Abrego, 2011; Gonzales, 2011; Hainmueller et al., 2017; Patler and Pirtle, 2018). Furthermore, undocumented workers also face large occupational barriers. They are prevented from working in occupations that require legal status (e.g. teachers, nurses, law enforcement) and shy away from professions that expose them to immigration enforcement, perhaps because they entail long-distance travel and frequent face-to-face interactions with the public or government officials (e.g. managers, secretaries and salespersons). In this sense, our work also relates to the literature on occupational licensing. Brucker et al. (2021) show that the formal recognition of foreign qualifications greatly improves immigrants' labor market outcomes in Germany, leading to full convergence to the earnings of their native counterparts. Kleiner and Krueger (2013) also document that licensing is associated with higher wages in the US, and Kleiner and Vorotnikov (2017) argue that relaxing licensing constraints amounts to a reduction in occupational barriers that leads to lower prices and higher consumer welfare. In a recent study, Blair and Chung (2017) have argued that occupational licensing can be a powerful tool to reduce the wage gaps of women and blacks (relative to white men) by reducing information asymmetries regarding worker productivity.

### **3** Theoretical Framework

Consider an economy with two occupations, indexed by o = 1, 2. Workers are heterogeneous in their idiosyncratic productivity vector  $\varepsilon = (\varepsilon_1, \varepsilon_2)$ , drawn from joint distribution  $f(\varepsilon_1, \varepsilon_2)$  with domain  $\mathbb{R}^2$  and assumed strictly positive over its domain. We will also refer to idiosyncratic productivity as *ability*. There are also two types of workers: documented (which includes natives) and undocumented (d = D, U). The measure of documented workers is normalized to 1 and the measure of undocumented workers is  $u \leq 1$ .

#### 3.1 The occupational choices of documented workers

For documented workers, wages are a function of productivity and each worker chooses the wage-maximizing occupation. As in the Borjas (1987) version of the Roy (1951) model, log wages are given by

$$\omega_{io} = \mu_o + \varepsilon_{io},\tag{1}$$

where  $\mu_o$  is the occupation-specific mean and  $\varepsilon_{io}$  the idiosyncratic productivity of worker i in occupation o. We will assume that occupation 2 has (weakly) higher average wages:  $\mu_2 \ge \mu_1$ . Note also that Equation (1) refers to log wages and, thus, negative values are meaningful.

Each individual *i* faces a vector of *potential* wages  $(\omega_{i1}, \omega_{i2})$ . But her *actual* wage depends on the chosen occupation. Thus, individuals' **optimal choice** is summarized

by the rule: choose occupation o = 2 if and only if

$$\omega_{i2} \ge \omega_{i1}$$
$$\mu_2 + \varepsilon_{i2} \ge \mu_1 + \varepsilon_{i1}$$
$$\varepsilon_{i2} - \varepsilon_{i1} \ge \mu_1 - \mu_2.$$

Hence, individuals self-select into the occupation that gets them the highest earnings (productivity).

The **optimal allocation** of workers to occupations is as follows. Let  $D_o$  denote the set of types (for documented workers) that choose occupation o. Figure 1 provides a graphical representation in the  $(\varepsilon_1, \varepsilon_2)$ -space. The solid line  $\varepsilon_2 = \varepsilon_1 - (\mu_2 - \mu_1)$  partitions the type space. Above this line documented workers choose occupation 2  $(D_2)$ . Below this line, documented workers choose occupation 1  $(D_1)$ . Thus

$$D_1 = \{ (\varepsilon_1, \varepsilon_2) : \varepsilon_2 < \varepsilon_1 - (\mu_2 - \mu_1) \}$$
$$D_2 = \{ (\varepsilon_1, \varepsilon_2) : \varepsilon_2 \ge \varepsilon_1 - (\mu_2 - \mu_1) \}.$$

As a result, (log) wages vary at the individual level within and across occupations:

$$\omega_{i} = \begin{cases} \mu_{1} + \varepsilon_{i1} & if \quad (\varepsilon_{i1}, \varepsilon_{i2}) \in D_{1} \\ \mu_{2} + \varepsilon_{i2} & if \quad (\varepsilon_{i1}, \varepsilon_{i2}) \in D_{2}, \end{cases}$$

where set  $D_1$  ( $D_2$ ) contains the documented workers that choose occupation 1 (occupation 2).

**Employment levels** can be computed integrating over the appropriate support of the density of types for documented workers.

$$Emp_1^D = Pr(D_1) = \int_{\varepsilon_1 = -\infty}^{\varepsilon_1 = \infty} \int_{\varepsilon_2 = -\infty}^{\varepsilon_2 = \varepsilon_1 - (\mu_2 - \mu_1)} f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1$$
(2)

and

$$Emp_2^D = Pr(D_2) = \int_{\varepsilon_1 = -\infty}^{\varepsilon_1 = \infty} \int_{\varepsilon_2 = \varepsilon_1 - (\mu_2 - \mu_1)}^{\varepsilon_2 = \infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1,$$
(3)

where  $Emp_{1}^{D} + Emp_{2}^{D} = 1$ .

#### **3.2** The occupational choices of undocumented workers

The idiosyncratic productivity distribution (ability) for undocumented is assumed to be identical to the one for documented workers:  $f(\varepsilon_1, \varepsilon_2)$ . Naturally, undocumented workers differ substantially from documented workers in terms of educational attainment and other productivity-related characteristics. For this reason, our empirical model will condition on observable characteristics.<sup>2</sup>

Importantly, in some occupations, undocumented workers cannot carry out the whole bundle of tasks due to lacking legal status. For instance, these workers cannot obtain a driver's license, travel by plane, obtain an occupational license, or freely interact with customers and government officials without the risk of apprehension. These limitations reduce the *productivity* of undocumented workers in these occupations, which in our setup are assembled under *occupation 2*. We note that, in contrast to a setup with purely psychic costs, here the employer pays a lower wage to an undocumented worker (in occupation 2) because of her diminished productivity relative to a documented worker in the same occupation and with the same idiosyncratic productivity. Because workers choose occupations, this **productivity penalty**, denoted by  $\phi \geq 0$ , amounts to an implicit *entry barrier* into occupation 2 that distorts the occupational choices of undocumented workers.

Additionally, undocumented workers are subject to **exploitation** by their employers, captured by parameter  $\tau \geq 0$ . As a result, undocumented workers' wages  $(\omega_{io}^U)$  are lower than their corresponding productivity by a factor  $\tau$ , assumed to be the same in both occupations.

In sum, depending on their choice of occupation, the wages of undocumented workers will be given by:

$$\omega_i^U = \begin{cases} (\mu_1 + \varepsilon_{i1}) - \tau & if \quad (\varepsilon_{i1}, \varepsilon_{i2}) \in U_1 \\ (\mu_2 + \varepsilon_{i2} - \phi) - \tau & if \quad (\varepsilon_{i1}, \varepsilon_{i2}) \in U_2 \end{cases}$$

where set  $U_1$  ( $U_2$ ) contains the undocumented workers that choose occupation 1 (occu-

<sup>&</sup>lt;sup>2</sup>Assuming that distribution  $f(\varepsilon_1, \varepsilon_2)$  is the same regardless of documentation status is convenient to derive closed-form solutions. Future work should consider relaxing this assumption, particularly in light of the recent findings in Patt et al. (2020) who find that Mexican migrants to the United States differ systematically from natives in terms of their manual (relative to cognitive) skills.

pation 2). The occupation-specific wages above make clear that the wage gap between equally skilled documented and undocumented workers in occupation 1 (in terms of  $\varepsilon_{i1}$ ) will be due exclusively to exploitation. In contrast, the wage gap in occupation 2 will also reflect the *productivity penalty* associated with lack of legal status.

Because of the productivity loss associated with lack of legal status, the occupational choices of undocumented immigrants will be distorted, leading to under-representation in some occupations and over-representation in others on the basis of the task bundle of each occupation. Namely, the **utility-maximizing occupation choice** for undocumented workers is to choose occupation o = 2 if and only if:

$$\omega_{i2}^U \ge \omega_{i1}^U$$
$$\mu_2 + \varepsilon_{i2} - \phi - \tau \ge \mu_1 + \varepsilon_{i1} - \tau$$
$$(\varepsilon_{i2} - \varepsilon_{i1}) \ge \phi - (\mu_2 - \mu_1),$$

which does not depend on the degree of employer exploitation  $\tau$ . Figure 2 summarizes the occupational choices of undocumented workers. In the type space, the indifference line for undocumented workers  $\varepsilon_2 = \varepsilon_1 + \phi - (\mu_2 - \mu_1)$  clearly lies higher up than the corresponding line for documented workers ( $\varepsilon_2 = \varepsilon_1 - (\mu_2 - \mu_1)$ ). The resulting **allocation** of undocumented workers to occupations can be described as follows. Let  $U_o$  denote the set of undocumented types that choose occupation o = 1, 2. Then

$$U_1 = \{(\varepsilon_1, \varepsilon_2) : \varepsilon_2 < \varepsilon_1 + \phi - (\mu_2 - \mu_1)\}$$

$$(4)$$

$$U_2 = \{(\varepsilon_1, \varepsilon_2) : \varepsilon_2 \ge \varepsilon_1 + \phi - (\mu_2 - \mu_1)\}.$$
(5)

#### 3.3 Occupational mismatch

The undocumented productivity penalty in occupation 2 induces occupational mismatch among a subset of undocumented workers: some of them would find it more beneficial to choose occupation 2 in the absence of the penalty, but instead inefficiently choose occupation 1. These worker types are only slightly better at occupation 2 and do not find it worthwhile to "pay" the cost to enter that occupation. Put otherwise, identical individuals make different occupational choices purely on the basis of documentation status. Accordingly, the **mismatch region** is given by

$$MM(\phi) = \{(\varepsilon_1, \varepsilon_2) : \varepsilon_1 - (\mu_2 - \mu_1) < \varepsilon_2 < \varepsilon_1 + \phi - (\mu_2 - \mu_1)\}.$$
(6)

We note that function MM maps values of  $\phi$  into subsets of the type space. Similarly,  $Pr(MM(\phi))$  is a function mapping values of  $\phi$  into the unit interval [0, 1]. It is easy to show that Pr(MM(0)) = 0 and  $Pr(MM(\phi))$  is increasing in  $\phi$ . Hence, the size of the mismatch region uniquely identifies the value of  $\phi$  (provided that there is positive density over the full domain of the density function).

Importantly, the occupation allocations of the two types of workers are related as follows:  $U_1 = D_1 \cup MM$  and  $D_2 = U_2 \cup MM$ . Thus the set of undocumented workers in occupation 1 equals the set of documented workers in that same occupation together with the mismatch set. Similarly, the set of documented workers in occupation 2 equals the union of the sets of undocumented workers in that same occupation and the set of missing undocumented workers who were mismatched into occupation 1. These relationships will be useful below when we compare the wage gaps between documented and undocumented workers.

#### **3.4** Gaps in employment shares

The employment levels of undocumented workers in each occupation can be defined by simply adapting the corresponding expressions for documented workers (given in Equation (2) and Equation (3)), and keeping in mind that  $Emp^U = Emp_1^U + Emp_2^U = u$ . Thus, overall employment in the economy in occupation o is simply given by

$$Emp_o = Pr(D_o) + uPr(U_o).$$

**Occupation shares.** One of the goals of the empirical analysis will be the comparison of the employment distributions for documented and undocumented workers. Specifically, we will focus on the undocumented-documented gap in occupational shares as follows:

$$egap_1 = \frac{Emp_1^U}{Emp^U} - \frac{Emp_1^D}{Emp^D} = Pr(U_1) - Pr(D_1)$$
(7)

$$egap_2 = \frac{Emp_2^U}{Emp^U} - \frac{Emp_2^D}{Emp^D} = Pr(U_2) - Pr(D_2).$$
 (8)

Clearly, due to the undocumented productivity penalty in occupation 2, undocumented workers will be under-represented in this occupation  $(egap_2 < 0)$  but over-represented in occupation 1  $(egap_1 > 0)$ . This follows easily from the earlier observations:  $Pr(U_1) =$   $Pr(D_1) + Pr(MM)$  and  $Pr(D_2) = Pr(U_2) + Pr(MM)$ . The following proposition characterizes the gaps in occupational shares by legal status:

**Proposition 1** When all workers choose their utility maximizing occupations,

 The employment share in occupation 2 is lower for undocumented workers than for documented workers, and the gap is increasing (in absolute value) in productivity penalty φ. In particular, the undocumented-documented gap is given by

$$egap_2 = -Pr(MM(\phi)) < 0.$$

2. Conversely, the employment share in occupation 1 is higher for undocumented workers than for documented workers and the gap is increasing in  $\phi$  and given by

$$egap_1 = Pr(MM(\phi)) > 0.$$

#### **Proof:** See Appendix B. ■

This finding will play an important role in the empirical analysis of the paper. Specifically, it shows that the comparison of the occupation shares of documented and undocumented workers identifies which occupations penalize undocumented workers' productivity, operating as entry barriers.

#### 3.5 Gaps in productivity and wages

**Wage schedules.** As a result of their occupational choices, the wage schedule for undocumented workers is given by:

$$\omega_{i}^{U} = \begin{cases} \mu_{1} + \varepsilon_{i1} - \tau & if \quad \varepsilon_{2} < \varepsilon_{1} - (\mu_{2} - \mu_{1}) \\ \mu_{1} + \varepsilon_{i1} - \tau & if \quad \varepsilon_{1} - (\mu_{2} - \mu_{1}) < \varepsilon_{2} < \varepsilon_{1} + \phi - (\mu_{2} - \mu_{1}) \\ \mu_{2} + \varepsilon_{i2} - \tau - \phi & if \quad \varepsilon_{2} > \varepsilon_{1} + \phi - (\mu_{2} - \mu_{1}). \end{cases}$$
(9)

It is helpful to describe the wage schedule for documented workers using the same

partition of the type space, namely,

$$\omega_{i}^{D} = \begin{cases} \mu_{1} + \varepsilon_{i1} & if \quad \varepsilon_{2} < \varepsilon_{1} - (\mu_{2} - \mu_{1}) \\ \mu_{2} + \varepsilon_{i2} & if \quad \varepsilon_{1} - (\mu_{2} - \mu_{1}) < \varepsilon_{2} < \varepsilon_{1} + \phi - (\mu_{2} - \mu_{1}) \\ \mu_{2} + \varepsilon_{i2} & if \quad \varepsilon_{2} > \varepsilon_{1} + \phi - (\mu_{2} - \mu_{1}). \end{cases}$$
(10)

It is worth emphasizing that wages equal productivity for documented workers and productivity, which has a common component  $(\mu_o)$  and an idiosyncratic one  $(\varepsilon_{io})$ . In contrast, there is a gap between the wages and productivity of undocumented workers due to the exploitation wedge  $(\tau)$ . Additionally, the productivity of undocumented workers employed in occupation 2 is diminished by a factor  $\phi$ , which captures the tasks involved in the occupation that cannot be fulfilled by individuals lacking legal status. This productivity penalty becomes an implicit barrier to the entry of undocumented workers into occupation 2.

Below we explore in detail the gaps in ability, productivity and wages between documented and undocumented workers *within* each occupation.

Self-selection in idiosyncratic productivity (ability). Because occupational choices are endogenous, and ability is one of the factors determining productivity and wages, it is important to understand the implications of occupational sorting in terms of selection in ability.

Because of the implicit entry barriers into occupation 2 faced by undocumented workers, it is tempting to expect higher average ability among undocumented workers in this occupation (relative to documented workers). By the same token, one would expect negative selection (in terms of ability) among undocumented workers in occupation 1. As it turns out, this is not a general result. As we shall see below, it requires distributional assumptions.

Let us begin by defining precisely what we mean by selection. We shall say that undocumented workers are *positively selected* (in ability) in occupation o = 1, 2 when

$$s_o(\phi) = E(\varepsilon_o|U_o) - E(\varepsilon_o|D_o) > 0.$$
(11)

Conversely,  $s_o(\phi) < 0$  indicates that undocumented workers are *negatively selected* in the occupation.

As noted, without further assumptions, we cannot rank the average ability of undoc-

umented and documented workers in either of the two occupations. This also implies that the average productivity and wages of the two types of workers within an occupation cannot be ranked either. Appendix A contains a simple numerical example illustrating that undocumented workers can be positively or negatively selected (in ability) in either occupation (or in both).

The following proposition provides sufficient conditions for the intuitive selection pattern to hold, where undocumented workers are positively selected in occupation 2 and negatively selected in occupation 1.

**Proposition 2** (Selection in ability). Let  $\phi > 0$  and suppose all workers choose the occupations that maximize their utility (wages).

1. A sufficient condition for positive selection of undocumented workers in occupation  $2 (s_2(\phi) > 0)$  is that,  $\forall \varepsilon_1$ , the conditional expectation of ability in occupation 2

$$E(\varepsilon_2|\varepsilon_1, \varepsilon_2 \ge \varepsilon_1 + \phi - (\mu_2 - \mu_1)) \tag{12}$$

is increasing in  $\phi$ . Moreover, the degree of positive selection in occupation 2 intensifies in  $\phi$ . So if  $\phi_1 > \phi_0 > 0$  then  $s_2(\phi_1) - s_2(\phi_0) > 0$ .

2. A sufficient condition for negative selection of undocumented workers in occupation 1  $(s_1(\phi) < 0)$  is that,  $\forall \varepsilon_2$ , the conditional expectation of ability in occupation 1

$$E(\varepsilon_1|\varepsilon_2,\varepsilon_1 > \varepsilon_2 - \phi + (\mu_2 - \mu_1)) \tag{13}$$

is decreasing in  $\phi$ . Moreover, the negative selection in occupation 1 intensifies in  $\phi$ . So if  $\phi_1 > \phi_0 > 0$  then  $s_1(\phi_1) - s_1(\phi_0) < 0$ .

**Proof:** See Appendix B.

As it turns out, the sufficient conditions in Proposition 2 hold for a large family of distributions. In particular, as stated in the next proposition, when idiosyncratic productivities are distributed uniformly or follow a bivariate normal distribution, undocumented workers will be positively selected in occupation 2 and negatively selected in occupation 1. However, the set of families for which the previous sufficient condition holds is much larger (Heckman and Honore, 1990).<sup>3</sup>

<sup>&</sup>lt;sup>3</sup>Heckman and Honore (1990) show that in the canonical Roy model, similar monotonicity conditions

**Proposition 3** (Special cases). Let  $\phi > 0$  and suppose all workers choose the occupations that maximize their utility (wages).

- 1. Let z > 0 and assume  $(\varepsilon_1, \varepsilon_2)$  is uniformly distributed with  $f(\varepsilon_1, \varepsilon_2) = 1/z^2$  for  $0 \le \varepsilon_1, \varepsilon_2 \le z$  and zero otherwise. Then  $s_1(\phi) < 0 < s_2(\phi)$ .
- 2. Let  $(\varepsilon_1, \varepsilon_2)$  follow a bivariate distribution (with correlation coefficient not necessarily equal to zero). Then  $s_1(\phi) < 0 < s_2(\phi)$ .

**Proof:** See Appendix B.

**Productivity and wage gaps.** It is helpful to define the undocumented-documented gaps in (log) productivity and wages within each occupation (o = 1, 2). Presumably, the average wage of documented workers is higher than the average wage among undocumented workers in the same occupation, and this gap is likely to be larger in occupation 2. Undocumented workers receive below-productivity wages (due to employer exploitation) and suffer a loss of productivity in occupation 2 due to the limitations arising from lack of legal status.

Let us begin by describing the undocumented-documented *productivity* gaps in occupation 2:

$$E(\mu_2 + \varepsilon_2 - \phi | U_2) - E(\mu_2 + \varepsilon_2 | D_2) = E(\varepsilon_2 | U_2) - E(\varepsilon_2 | D_2) - \phi$$
  
=  $s_2(\phi) - \phi.$  (14)

Let us now turn to the gap in (log) wages in the same occupation. Since wages equal productivity for documented workers, and differ from productivity by a factor  $\tau$  for undocumented workers, the undocumented-documented log wage gap reduces to

$$wgap_2 = E(\omega_2|U_2) - E(\omega_2|D_2)$$
  
=  $s_2(\phi) - \phi - \tau.$  (15)

It is worth noting that in situations where Proposition 2 and Proposition 3 apply and, therefore, undocumented workers are positively selected in occupation 2  $(s_2 > 0)$ , the endogenous occupational choices made by individuals *reduce* the productivity and wage gaps (in absolute value) between documented and undocumented workers.

as those invoked in Equation (12) and Equation (13) hold for log concave distributions, which includes the normal, uniform, beta and extreme value distributions.

The expressions for the undocumented-documented gaps in productivity and wages for occupation 1 are slightly simpler because there is no productivity loss for workers lacking legal status in this occupation. Respectively, the productivity and log wage gaps are given by

$$E(\mu_1 + \varepsilon_1 | U_1) - E(\mu_1 + \varepsilon_1 | D_1) = s_1(\phi)$$
  

$$wgap_1 = s_1(\phi) - \tau.$$
(16)

The following proposition gathers the expressions for the log wage gaps in each occupation, which will play a central role in our empirical application.

**Proposition 4** (Log wage gaps). Let  $\phi > 0$  and suppose all workers choose the occupation that maximizes their utility (wages).

1. The occupation-specific undocumented-documented wage gaps are given by

$$wgap_1 = E(\omega_1|U_1) - E(\omega_1|D_1) = -\tau + s_1(\phi)$$
(17)

$$wgap_2 = E(\omega_2|U_2) - E(\omega_2|D_2) = -\phi - \tau + s_2(\phi).$$
(18)

- 2. The average wage of undocumented workers in occupation 1 will be higher than the corresponding value for documented workers ( $wgap_1 > 0$ ) if and only if  $s_1(\phi) > \tau$ .
- 3. The average wage of undocumented workers in occupation 2 will be higher than the corresponding value for documented workers ( $wgap_2 > 0$ ) if and only if  $s_2(\phi) > \phi + \tau$ .

**Proof:** Follows trivially from Equation (15) and Equation (16).  $\blacksquare$ 

For reasons that will become clear shortly, it is helpful to consider the difference in wage gaps:

$$wgap_1 - wgap_2 = \phi + (s_1 - s_2), \tag{19}$$

where the exploitation wedge  $\tau$  cancelled out.

**Empirical prediction.** Next, we show how the previous results can be used to empirically estimate a lower bound for the productivity penalty emerging from lack of legal status. It is helpful to consider first the scenario **without** endogenous occupational

sorting. In this case, both the exploitation wedge parameter  $\tau$  and the productivity penalty parameter  $\phi$  can be identified using the appropriate undocumented-documented wage gaps. It follows from Equation (17) and Equation (19) that

$$\tau = -wgap_1 = E(\omega_1|D_1) - E(\omega_1|U_1)$$
  
$$\phi = wgap_1 - wgap_2.$$

In words, in the absence of endogenous sorting, the degree of wage exploitation can be gauged from the wage gap in occupation 1, and the productivity penalty from the double difference in log wages  $(wgap_1 - wgap_2)$ .

In reality, occupational sorting *cannot* be ruled out. As became clear in Proposition 4, this will affect the wage gaps between documented and undocumented workers and also complicates the identification of parameters  $\tau$  and  $\phi$ . Nonetheless, as stated in the next proposition, a lower bound for the undocumented productivity penalty can still be estimated under fairly reasonable distributional assumptions.

**Proposition 5** (Lower bound for productivity penalty). Assume the joint ability distribution, f, is such that undocumented workers are positively selected in occupation  $2 (s_2(\phi) > 0)$  and negatively selected in occupation  $1 (s_1(\phi) < 0)$ . Then

$$\phi > wgap_1 - wgap_2. \tag{20}$$

**Proof:** The proof follows trivially from Equation (19) and assumptions  $s_1 < 0 < s_2$ . Simply,  $wgap_1 - wgap_2 = \phi + (s_1 - s_2) > \phi$ .

In words, the difference in wage gaps (in turn, a double difference in wages across legal status and occupations) provides a lower bound for the value of  $\phi$ . To fix ideas, consider the following example. Suppose that  $wgap_1 = w_1^U - w_1^D = -0.05$  and  $wgap_2 = w_2^U - w_2^D = -0.10$ . Then  $\phi > 0.05$ , that is, lack of legal status entails a productivity loss equal to 5 log points or larger.

It is important to keep in mind that  $wgap_o$  has been defined as the undocumenteddocumented wage gap, and is likely to be negative for both occupations. To be informative, the lower bound in Equation (20) requires that the undocumented-documented wage gap in occupation 1 be *larger* than in occupation 2. It follows from Equation (17) and Equation (18) that this will be the case provided that the undocumented productivity penalty  $\phi$  overpowers the effects of ability selection. Namely, the *productivity penalty*  lower bound  $(wgap_1 - wgap_2)$  will be positive provided that  $\phi > s_2(\phi) - s_1(\phi)$ , which depends on distributional assumptions and is ultimately an empirical question.<sup>4</sup>

#### 3.6 The Effects of Legalization

Granting legal status to undocumented workers eliminates the frictions ( $\tau = \phi = 0$ ). The following proposition gathers all the effects of such a measure in the context of our theory:

#### **Proposition 6** (Legalization). As a result of legalization ( $\phi = \tau = 0$ ), we expect:

- 1. Occupational switch: the previously mismatched undocumented workers switch from occupation 1 to occupation 2, and the gaps in employment shares vanish.
- 2. The productivity and wages of workers with identical idiosyncratic productivity types converge, that is, the wage schedule for all types is given by Equation (10).
- 3. Wages increase for all undocumented workers according to the following schedule:

$$\Delta \omega^{U} = \begin{cases} \tau & if \quad \varepsilon_{2} < \varepsilon_{1} - (\mu_{2} - \mu_{1}) \\ \tau + (\mu_{2} - \mu_{1}) + (\varepsilon_{2} - \varepsilon_{1}) & if \quad \varepsilon_{1} - (\mu_{2} - \mu_{1}) < \varepsilon_{2} < \varepsilon_{1} - (\mu_{2} - \mu_{1}) + \phi \\ \tau + \phi & if \quad \varepsilon_{2} > \varepsilon_{1} - (\mu_{2} - \mu_{1}) + \phi. \end{cases}$$

4.  $\Delta \omega^U(\varepsilon_2)$  is a (weakly) increasing and continuous function in  $\varepsilon_2$ . Namely, the higher  $\varepsilon_2$  individuals experience larger wage gains.

The statements in Proposition 6 follow easily from the earlier results. The elimination of the undocumented-documented gaps in employment shares across occupations follows from the vanishing of the mismatch area (i.e. Pr(MM(0)) = 0)). Next, setting  $\tau = \phi = 0$  implies that the wage schedule for undocumented workers collapses to the schedule for documented workers (Equation (10)).

Claim 3 in the proposition simply follows from recognizing that the post-legalization wage schedule for previously undocumented workers (Equation (9)) is the same as the wage schedule for documented workers (Equation (10)) so that  $\Delta \omega^U = \omega^D - \omega^U$ . While theoretically intuitive, it is comforting to see that there is widespread empirical support for the prediction of generalized wage gains for undocumented workers following a

<sup>&</sup>lt;sup>4</sup>As shown in Section C.1, the productivity penalty lower bound is informative for the uniform distribution over the range of plausible values for  $\phi$ .

legalization process, both in the United States (Rivera-Batiz, 1999; Kossoudji and Cobb-Clark, 2002; Amuedo-Dorantes et al., 2007; Lozano and Sorensen, 2011; Pan, 2012) and abroad (Monras et al., 2017).

More interestingly, claim 4 states that the wage increases experienced by undocumented workers are heterogeneous in size depending on their relative ability in occupation 2 ( $\varepsilon_2 - \varepsilon_1$ ). As illustrated in Figure 3, we can partition undocumented workers into three groups on the basis of their occupation-2 ability,  $\varepsilon_2$ , given their occupation-1 ability. Individuals with the lowest values of  $\varepsilon_2$  were initially employed in occupation 1 and remain so after legalization. These workers experience a wage increase of  $\tau$  (but no increase in productivity). Undocumented workers with higher values of  $\varepsilon_2$  all experience increases in productivity, but to a different degree.

Undocumented individuals with the *highest* values of  $\varepsilon_2$  were initially employed in occupation 2 and remain in the same occupation after legalization. These workers experience a productivity increase given by  $\phi$  and a wage increase of  $\tau + \phi$ , which turns out to be the largest increase experienced by any undocumented worker. In turn, those with intermediate values of  $\varepsilon_2$  switch occupations when they receive legal status and, as a result, see their wage increase by  $\tau + (\mu_2 - \mu_1) + (\varepsilon_2 - \varepsilon_1)$ . Clearly, switchers with higher values of  $\varepsilon_2$  enjoy larger gains. It is also easy to show that the upper bound for the wage increase experienced by switchers is  $\tau + \phi$ .<sup>5</sup> Thus proving that the overall wage gain is a (weakly) increasing function of  $\varepsilon_2$ .

**GDP increase**. Proposition 6 has an important corollary in regards to the net output effects of legalization. Legalization increases aggregate output (defined as the integral of individual productivity) in the economy. Individual productivity gains also vary across individuals according to the schedule  $\Delta \omega^U(\varepsilon_2) - \tau$ , which is also (weakly) increasing in  $\varepsilon_2$  (as in Figure 3). Thus, the increases in worker productivity arise for two reasons: previously undocumented workers are now matched to their optimal frictionless occupations, and the productivity penalty disappears for those already working in occupation 2 prior to legalization.<sup>6</sup>

It is straightforward to derive an expression for the increase in (the log of) GDP (denoted by  $g^Y$ ) by adding up all the individual productivity increases (and recalling

<sup>&</sup>lt;sup>5</sup>Similarly, the associated productivity increase is given by  $(\mu_2 - \mu_1) + (\varepsilon_2 - \varepsilon_1)$ . Continuity in  $\varepsilon_2$  can be easily checked by evaluating the expression  $\tau + (\mu_2 - \mu_1) + (\varepsilon_2 - \varepsilon_1)$  at the cutoff values  $\varepsilon_2 = \varepsilon_1 + (\mu_1 - \mu_2)$  and  $\varepsilon_2 = \varepsilon_1 + (\mu_1 - \mu_2) + \phi$ .

<sup>&</sup>lt;sup>6</sup>Quantification of the actual increase in GDP will also depend on the price of the output produced in each occupation, which presumably produce different goods.

that the measure of undocumented workers in the economy is  $u \leq 1$ ):<sup>7</sup>

$$g^{Y}(\phi) = u \int_{\varepsilon_{1}=-\infty}^{\varepsilon_{1}=\infty} \int_{\varepsilon_{2}=-\infty}^{\varepsilon_{2}=\infty} \left[ \Delta \omega^{U}(\varepsilon_{1},\varepsilon_{2}) - \tau \right] f(\varepsilon_{1},\varepsilon_{2}) d\varepsilon_{1} d\varepsilon_{2}$$

$$= u \left( \phi Pr(U_{2}) + (\mu_{2} - \mu_{1}) Pr(MM) + \int_{-\infty}^{\infty} \int_{\varepsilon_{1}-(\mu_{2}-\mu_{1})}^{\phi+\varepsilon_{1}-(\mu_{2}-\mu_{1})} (\varepsilon_{2} - \varepsilon_{1}) f(\varepsilon_{1},\varepsilon_{2}) d\varepsilon_{1} d\varepsilon_{2} \right)$$

$$(21)$$

The first term in the parenthesis corresponds to the GDP gains arising from the removal of the productivity loss for undocumented workers that were already employed in occupation 2 even before gaining legal status. The second and third terms are the productivity gains obtained by the undocumented workers that were mismatched prior to legalization. They now enter an occupation with higher average productivity (since  $\mu_2 \ge \mu_1$ ) and, in addition, they have comparative advantage in the new occupation ( $\varepsilon_2 > \varepsilon_1$ ). Clearly, we are adding three positive terms and thus GDP will unambiguously increase. Quantifying this expression requires distributional assumptions on the joint distribution of ability and an estimate of undocumented productivity penalty  $\phi$ .<sup>8</sup>

Last, it is worth emphasizing that in the absence of an explicit undocumented productivity penalty (i.e. with  $\phi = 0$  prior to legalization), legalization would still raise the wages of all undocumented workers. Obviously, undocumented workers would not be subject to exploitation any more, which would raise everyone's wage by roughly  $\tau$  percent. The wage increase would be identical for all workers (and equal to  $\tau$ ). In this case, no undocumented workers were mismatched prior to the change in legal status. Therefore, legalization would not change the productivity of any undocumented workers and, as a result, it would not entail a net increase in GDP (or overall tax revenue). In terms of Equation 3.6,  $\Delta GPD(\phi = 0) = 0$ . In sum, legalization would simply redistribute income from employers toward undocumented workers.

The next section presents the data we will use to compute the undocumenteddocumented gaps in occupational shares and wages, which provide the essential inputs to produce a lower bound for the productivity penalty from lack of legal status. Section 6.4 will revisit to produce a back-of-the-envelope estimation of the GDP gains from legalization for the United States.

<sup>&</sup>lt;sup>7</sup>Since the productivity terms throughout the model are in logs, their aggregation in Equation 3.6 can be interpreted as the percent change in GDP.

 $<sup>^{8} {\</sup>rm Section} \ {\rm C.2}$  provides a closed form solution for the uniform distribution.

## 4 Data and Summary Statistics

We use a special extract of the American Community Survey provided by the Center for Migration Studies (2014). Besides the usual information on employment, skills and wages, this confidential dataset contains a sophisticated imputation for documentation status developed by Warren (2014). These data have been used to estimate, by means of calibration and simulation methods, the economic contribution of undocumented workers (Edwards and Ortega, 2017) and the consequences of providing legal status to Dreamers (Ortega et al., 2019).

Even though several authors have developed newer datasets containing similar imputations (e.g. Albert (2021) and Peri and Zaiour (2021) using the CPS and Borjas and Cassidy (2019) using the ACS), we restrict our analysis to years 2010-2012. The reason is that President Obama's *Deferred Action on Childhood Arrivals* was rolled out starting at the very end of 2012. This program provided beneficiaries with reprieve from deportation and two-year renewable work permits, which has been shown to have improved substantially the labor market outcomes of its recipients (Pope, 2016; Amuedo-Dorantes and Antman, 2017). Since we cannot identify DACA recipients in the data, it is preferable to restrict the analysis to the pre-DACA period.

The unauthorized status imputation was first proposed in the 1990's and many authors have contributed to their development over the last few decades (Passel and Clark, 1998; Baker and Rytina, 2013; Warren and Warren, 2013; Passel and Cohn, 2015; Warren, 2014). The procedure is a 2-step process: (1) applying 'logical edits' to identify legal residents on the basis of the information in the ACS; and (2) re-weighting individual observations to match official unauthorized population estimates by country of origin. The main logical edits rely on information on year of arrival (because of the 1986 IRCA amnesty), country of origin, occupation, industry, and receipt of government benefits.<sup>9</sup> Strictly speaking, we should refer to *likely* unauthorized individuals but, for simplicity, we will simply refer to unauthorized (or undocumented) individuals.

Our data (for year 2012) show that most undocumented have been present in the United States for 16 years or more, and some have resided in the country for three decades (Figure 4). As a result, undocumented immigrants are deeply rooted in their local communities. Furthermore, about a third of the undocumented (amounting to approximately 3 million individuals) were brought to the country as children (often

<sup>&</sup>lt;sup>9</sup>Warren (2014) argues that the imputation accounts for 89% of unauthorized residents, which increases to 93% if we add individuals that were unauthorized at some point in the past. Other studies assessing the validity of this methodology are Pastor and Scoggins (2016) and Van Hook et al. (2015).

known as *Dreamers*). We restrict the analysis to adult, full-time employees. In our data, we estimate that 5.1 million workers are likely undocumented, accounting for about 5% of full-time employment.

Let us now present descriptive statistics for the variables we will use in the estimation (Table 1). The (unweighted) data contain 2.7 million observations and about 4% correspond to likely undocumented workers. The rest are documented workers, which contains both US-born individuals and foreign-born with legal status. The mean hourly wage is \$23 across all workers (pooling both documented and undocumented). About 7% of the sample did not graduate from high-school and 35% obtained a 4-year college degree.<sup>10</sup> Last, the table also reports the continent of origin of each individual, the degree of English fluency and an indicator for having arrived in the country before the age of 10, which is meant to proxy for being educated in the United States. Specifically, 7% of the individuals in the sample were born in South or Central America (and 5% in Asia), 92% are fluent in English and 87% arrived in the United States by age 10.<sup>11</sup> The last column in the table reports the means of the variables for the undocumented subsample. Compared to all workers, on average undocumented workers are 7 years younger, 15 percentage-points less likely to be female, much more likely to lack a high-school degree (by 36 percentage-points) and earn lower hourly wages (by about \$9).<sup>12</sup>

## 5 Gaps in Employment Shares

The main goal of the empirical analysis in this paper is to estimate the lower bound for the undocumented productivity penalty. As we showed earlier, estimation requires computing the undocumented-documented wage gaps in occupations where lacking legal status lowers productivity and in occupations where this is not the case. The productivity penalty operates as an entry barrier into the corresponding occupations (Hsieh et al., 2019) because only individuals with a large comparative advantage (in terms of idiosyncratic productivity) in those occupations will find it worthwhile to enter.

The first step in measuring the productivity penalty is to partition occupations in

<sup>&</sup>lt;sup>10</sup>The ACS data provide 10 categories for the educational attainment of individual respondents. The lowest level is for individuals with completed education up to 4th grade, followed by individuals that completed up to 8th grade. The top two educational categories are a 4-year college degree, and having completed 5 or more years of college (including graduate studies). We aggregate these categories into high-school dropouts (HSD), high-school graduates (HSG) and College graduates (CoGrad).

<sup>&</sup>lt;sup>11</sup>US-born individuals are considered as 'having arrived' in the United States before age 10.

<sup>&</sup>lt;sup>12</sup>Peri and Zaiour (2021) report similar information for the 2020-2021 likely undocumented population in the United States.

two groups. This partition will be based on the comparison of the employment distributions of documented and undocumented workers (as in Proposition 1), after netting out observable characteristics. In particular, occupations were undocumented workers are under-represented relative to observationally equivalent documented workers will be identified as occupations displaying an undocumented productivity penalty (that is, occupation 2 in our theoretical framework). After we partition occupations in this manner, we will then use the average wage gaps between documented and undocumented workers to estimate the productivity penalty (or its lower bound).

The career choices of unauthorized workers are often shaped by explicit entry barriers, such as when an occupation requires a license that is only available for legal residents. However, entry barriers may also be implicit and stem from the tasks involved in a given occupation. Some occupations require wide exposure to the public, the ability to drive (legally), or to travel long distances, all of which increase the risk of apprehension faced by undocumented workers or are completely out of bounds for them (e.g. air travel).

#### 5.1 Unconditional gaps in employment shares

It is helpful to begin by comparing the employment distributions of documented and undocumented workers, disregarding for now differences in individual characteristics. Namely, we simply compute undocumented-documented gaps in employment shares for all (2-digit) occupations.

Unconditional employment (occupational) shares can be computed easily as the proportion of individuals employed in each occupation relative to the total number of fulltime employed individuals in the group. Accordingly, for each occupation o and group g = D, U, we compute  $p_o^g = \frac{Emp_o^g}{Emp^g}$ , where the numerator refers to individuals of group g in occupation o, and the denominator is the overall employment individuals of group g across all occupations. Then the unconditional gap in employment shares is simply  $egap_o = p_o^U - p_o^D$  (as in Equation (7) and Equation (8)). We note that the size of the gap reflects both the differences in the occupational distribution of the two groups and the overall size of the occupation in terms of employment (as well as differences in average characteristics between the two groups of workers).

Table 2 presents the 20 occupations with the largest gaps, which employ 12% of undocumented workers and 45% of documented workers. The top 5 occupations in the list are (23) Teachers and Instructors (up to secondary schooling), (32) Nurses and Therapists, (04) Managers and administrators, (57) Secretaries and (08) Accountants,

Auditors and Financial specialists.<sup>13</sup> Columns 1 and 2 report the estimated employment levels in each occupation for documented and undocumented workers, respectively. For instance, we estimate that roughly 4 million documented workers were employed as *Teachers and Instructors*, but only about 20,000 undocumented workers were employed in this occupation. As shown in columns 3 and 4, the corresponding employment shares also vary by group. *Teachers and Instructors* account for 4.4% of employment for documented workers, but only for 0.4% for undocumented workers, which results in a (negative) 4.1 percentage-point gap.<sup>14</sup> The second row shows a (negative) 2.7 percentagepoint gap among *Nurses and Therapists*, and a 2.3 percentage-point gap for *Managers and Administrators*. The bottom row in the table reports the total employment level and share for all occupations in which undocumented workers are under-represented (that is,  $egap_o < 0$ ). About 1.26 million undocumented workers are employed in this occupations (for 63.84 million documented workers). Importantly, these occupations only account for 24.7% of the employment of undocumented (compared to 70.8% for documented workers).

More generally, we note that many of the occupations in Table 2 require occupational licenses (e.g. teachers and healthcare professionals). However, we also find occupations that do not require licensing but entail face-to-face interactions with customers or government officials (e.g. secretaries, retail sales and clerks) or often require driving (e.g. salesmen or mail carriers). Last, we note that some of the occupations in the table entail college degrees, such as teachers, registered nurses or lawyers. Thus, differences in educational attainment and other individual characteristics are also partly responsible for the gaps in employment shares, but have nothing to do with idiosyncratic productivity differences. To obtain a cleaner identification of the occupations where undocumented workers are under-represented relative to *observationally equivalent* documented workers, we next extend our analysis to take into account individual characteristics in age, gender, education, and so on.

 $<sup>^{13}</sup>$ We estimate zero likely undocumented workers employed in *Law Enforcement* or belonging to the *Military*.

<sup>&</sup>lt;sup>14</sup>Until 2014 licensing requirements for teachers in all U.S. states required legal residence. Thus the likely undocumented workers employed in this occupation are a combination of foreign teachers on temporary visas (who are misclassified as undocumented by the imputation) and truly undocumented individuals working as instructors (in community organizations or after-school programs). A similar point was discussed in Borjas and Cassidy (2019). Since 2014, several states (such as California and New York) have adopted changes in licensure requirements to allow DACA recipients access to teaching occupations (Calvo, 2017).

#### 5.2 Conditional gaps in occupational shares

As we discussed earlier, undocumented workers are younger, less likely to be female and much less educated, on average, than documented workers (Table 1). These characteristics are likely to shape their occupational choices and obscure the pattern of selection in idiosyncratic ability highlighted in the theory section of the paper.

To account for differences in individual characteristics, we build *conditional* gaps in employment shares by estimating a series of occupation-specific binomial probit models. Specifically, let  $d_{io}$  denote an indicator function taking a value of one if individual *i* is employed in occupation *o*, and zero otherwise. Then we postulate that

$$Prob(d_{io} = 1|X_i) = \Phi(\alpha_o + \beta_o Undoc_i + X'_i \gamma_o), \qquad (22)$$

where  $\Phi$  is the CDF of the standard normal distribution. Dummy variable  $Undoc_i$  takes a value of one for likely undocumented individuals and zero otherwise. Thus,  $\beta_o < 0$ suggests that there exists an entry barrier into occupation o for undocumented workers, which we interpret as a productivity penalty. The vector of individual characteristics,  $X_i$ , includes dummies for age groups, gender, state of residence and educational categories. On the basis of the (maximum likelihood) estimates of the coefficients above,<sup>15</sup> we compute *conditional* average effects of undocumented status on employment shares:

$$E^{U}(d_{io}|\overline{X}) - E^{D}(d_{io}|\overline{X}) = \Phi(\hat{\alpha_{o}} + \hat{\beta_{o}} + \hat{\gamma_{o}}\overline{X}) - \Phi(\hat{\alpha_{o}} + \hat{\gamma_{o}}\overline{X}),$$
(23)

where  $E^g$  indicates that the expectation integrates over the subset of individuals belonging to group g = D, U. Importantly, we impose the same distribution of individual characteristics on documented and undocumented workers (equal to the overall sample mean) in order to neutralize the effect of differences in individual characteristics between the two groups.

Columns 1-6 in Table 3 compare the mean age, share of females and share of college graduates among documented and undocumented workers in each occupation. Typically, undocumented workers are younger and less likely to be female. In addition, in some occupations undocumented workers are less likely to be college graduates than documented workers (e.g. *Teachers & Instructors* or *Salespersons*), while the converse is true in other occupations (e.g. *Nurses & Therapists* or *Secretaries*).

Column 7 reports the *conditional* gaps in employment shares, computed as in Equa-

<sup>&</sup>lt;sup>15</sup>Sampling weights are used in the estimation of all models in the paper.

tion (23). The estimates show that the set of occupations where undocumented status entails a larger reduction in the probability of employment is similar to those based on unconditional gaps (Table 2). In particular, the occupation with the largest employment share gap is still the same (*Teachers & Instructors*) and 8 out of the 10 top occupations in Table 3 belong also to the top 10 in Table 2. However, the gaps are now smaller given that differences in individual characteristics have been neutralized. Specifically, while the unconditional employment share for *Teachers & Instructors* was 4.1 percentage points lower for undocumented workers, the corresponding conditional gap is only 2.4 percentage points.<sup>16</sup>

Summing up, our analysis has uncovered large differences in the occupational distributions of workers on the basis of their legal status that are not accounted for by differences in individual characteristics, indicating occupational mismatch due to (explicit or implicit) entry barriers. While in some cases these barriers are based on regulations, in others the barriers stem from the tasks required in those occupations. In light of the findings in Hsieh et al. (2019), the large disparities in the employment distributions of observationally similar documented and undocumented workers suggest that lack of legal status distorts the occupational choices of undocumented workers, leading to misallocation of talent and productivity losses.

## 6 Wage gaps and the undocumented productivity penalty

Next, we turn to estimate the wage gaps between observationally equivalent documented and undocumented workers, and to use these estimates to learn about the productivity loss associated with lack of legal status.

To bridge the gap between wages and productivity, we need to address two challenges. First, we need to adjust for differences in observable characteristics between documented and undocumented workers, as we did to estimate the conditional employ-

<sup>&</sup>lt;sup>16</sup>As we show in the Appendix (Table 6), the occupations with the largest employment share gaps remain largely unchanged when we use documented *foreign-born* workers as the comparison group. The top 3 occupations in this case also belong to the top 5 in Table 3, and the size of the employment share gaps are also similar, averaging 1.3 percentage points, compared to 1.6 percentage points in Table 3. We have also estimated the occupational choice probit models controlling for English fluency and arrival into the US before age 10. The resulting set of top 20 occupations by entry barriers is also fairly similar to Table 3. Specifically, 18 of the 20 occupations in Table 3 are also in the top 25 that results when adding the new controls.

ment share gaps. Besides the roles of education, age (potential experience) and other sociodemographic characteristics, labor economists have long recognized (Chiswick (1991), Chiswick et al. (2005)) that immigrants with an imperfect command of English will suffer a productivity and wage loss. The richness of the ACS allows us to build detailed controls to mitigate this problem.

The second challenge to identify the productivity penalty from lack of legal status entails accounting for employer exploitation and distortions in the occupational choices of undocumented workers. Our theoretical analysis delivered a strategy to estimate the undocumented productivity penalty (or a lower bound for it), based on the comparison between the undocumented-documented wage gaps in occupations with and without an undocumented productivity penalty. As stated in Proposition 5, the productivity penalty associated with lack of legal status (parameter  $\phi$  in our theoretical framework) will be at least as large as the difference between the undocumented-documented wage gap in occupations where lack of legal status does *not* entail a productivity penalty (*wgap*<sub>1</sub> in our earlier notation) and the corresponding wage gap in occupations where it does (*wgap*<sub>2</sub>). As we argued earlier, this difference in wage gaps is a function of the undocumented productivity penalty and the differences in selection (in terms of ability) among undocumented workers in the two sets of occupations. As we showed, under plausible conditions, (*wgap*<sub>1</sub> - *wgap*<sub>2</sub>) provides a *lower bound* for the undocumented productivity penalty.

Let us turn now to the empirical specification we shall employ to estimate the wage gaps between workers in occupations with and without undocumented productivity penalties. We postulate that the log hourly wage for individual i employed in occupation o is given by

$$\ln w_{io} = \alpha_o + \beta^U Undoc_i + \gamma Undoc_i \times Barrier_o + X'_i \Lambda + u_{io}, \tag{24}$$

where  $\alpha_o$  are occupation fixed-effects,  $Undoc_i$  is an indicator for (likely) undocumented status,  $Barrier_o$  identifies occupations with barriers to the entry of undocumented workers. This dummy variable is based on the *conditional* employment share gaps between documented and undocumented workers estimates in Section 5.2. Last,  $X_i$  is a vector of observable characteristics, including continent of origin dummies, an indicator for English fluency (taking a value of one for US-born workers and for foreign-born individuals reporting high fluency), and an indicator for having arrived in the country before the age of 10 (Bleakley and Chin, 2010). Coefficient  $\beta^U$  identifies the log wage gap associated with lack of legal status in occupations without significant entry barriers for undocumented workers. We expect this coefficient to be negative in light of the evidence of wage exploitation in previous studies. In terms of the notation introduced in our theory section,  $\beta^U = wgap_1$ . Accordingly, the corresponding log wage gap in occupations with barriers due to diminished productivity for undocumented workers will be given by  $wgap_2 = \beta^U + \gamma$ . Therefore, the lower bound for the undocumented productivity penalty (derived in Proposition 5) is given by the double difference

$$wgap_1 - wgap_2 = -\gamma. \tag{25}$$

Before turning to estimation, there is an important implementation decision regarding the estimation sample. To estimate the wage gaps associated with lack of legal status, using foreign-born documented workers as benchmark is the most natural choice. In other words, one could restrict the estimation sample to foreign-born workers only, as done in Borjas and Cassidy (2019), Albert (2021) and Peri and Zaiour (2021). However, this choice entails some shortcomings. First of all, restricting to foreign-born workers only entails a drastic reduction in sample size (of 83%, from 1.8 to 0.3 million observations), which might result in many 2-digit occupations with very few observations.

In addition, it is worth keeping in mind that undocumented status is not directly observed, but *imputed* and thus subject to measurement error. This measurement error has a particular structure that probably leads to an underestimation of the wage gap in occupations with entry barriers. In some of these occupations (e.g. teachers), we find individuals labelled as likely undocumented even though they are legally authorized to work. For instance, this may be the case for temporary visa holders. The wages for these individuals are likely to be in line with the wages of comparable documented workers because they will not be subject to employer exploitation and will not suffer a loss in productivity. As a result, we will underestimate the wage gap between documented and likely undocumented workers in occupations where this measurement error is more prevalent. To the extent that this is the case in occupations with explicit entry barriers, such as nurses or teachers, we would be obtaining a downwardly biased estimate of the lower bound for  $\phi$  on the foreign-born sample, which would be less informative. This is why our preferred estimates are based on the full sample, where the wages of undocumented workers are benchmarked using all observationally similar documented workers (on the basis of a model that controls for origin continent, English fluency and age of arrival in the United States besides other socio-demographic characteristics). At

any rate, we shall report the estimates obtained using both estimation samples.

#### 6.1 Baseline estimates

Table 4 collects the estimates for the linear wage model described in Equation (24). The top panel reports estimates based on the full sample, including natives, and the bottom panel focuses on foreign-born workers only. The table reports two coefficients: the estimate corresponding to the undocumented status indicator ( $\beta^U$ ) and its interaction with the dummy identifying occupations with large entry barriers ( $\gamma$ ), which we define as the 20 occupations with the largest conditional employment share gaps in Equation (23). Both samples are restricted to full-time employed individuals.

As seen in column 1 (top panel) of Table 4, the raw hourly wage for undocumented workers is about 40 log points lower than for documented workers in occupations without significant entry barriers to undocumented workers, consistent with the value reported in Borjas and Cassidy (2019) (also based on ACS data). Controlling for state of residence, age and gender hardly affects this gap. However, controlling for educational attainment reduces the gap to 18 log points. If we also account for continent of origin, English fluency and arriving in the country as a child, the gap shrinks down to 3 log points (column 5), which underscores the importance of including detailed controls for the determinants of worker productivity. For our purposes, the main specification is reported in column 6, which includes 2-digit occupation dummies and, therefore, accounts for occupationspecific differences in wages. The estimated coefficient for the undocumented indicator shows that undocumented workers employed in occupations without entry barriers have hourly wages that are only 2 log-points lower than observationally similar documented workers in those same occupations. Interpreted through the lens of Equation (17), this estimate suggests that the degree of employer exploitation in the average occupation  $(\tau)$ is small.

Let us now turn to the estimation of the double difference in wages that identifies the lower bound for the undocumented productivity penalty. Jumping directly to column 5 (which includes all control variables), we estimate that the average productivity loss suffered by undocumented workers employed in occupations with entry barriers is 12 log points. The next column reports estimates for a specification that includes (2-digit) occupation dummies, which accounts for the fact that some occupations have generally higher wage levels than others. The estimated lower bound for the undocumented productivity penalty remains unaltered at 12 log points. In sum, these estimates indicate that the undocumented productivity penalty is sizable and legalization would have quantitatively important effects on the productivity and wages of many undocumented workers.<sup>17</sup>

#### 6.2 Robustness

Next, let us examine the robustness of our main finding. Our baseline estimation of the lower bound relied on considering only the 20 occupations with the largest conditional employment share gaps (as in Equation (23)). Instead, column 7 considers a more expansive definition that includes the top 30 occupations with the largest employment gaps. As expected, the double difference in wages diminishes (in absolute value) because the two sets of occupations are now less apart in terms of the endogenous occupational sorting of undocumented workers. However, the lower bound for the undocumented productivity penalty remains large (at roughly 9%).

Let us now turn to analyze the robustness of our estimates to restricting the estimation sample to foreign-born individuals. The estimates are reported in the bottom panel of Table 4. Moving directly to columns 5 and 6, we estimate the lower bound for the undocumented productivity penalty to be lower (as expected on the basis of the earlier discussion regarding measurement error). However, the lower bound remains large (estimated at 8 log points).

It is also worth comparing our estimates to those reported in previous studies. Using ACS data (for years 2010-2012), Borjas and Cassidy (2019) report the *conditional* wage gap to range between 5 and 8 percent. In turn, using CPS data (for years 1994-2016), Albert (2021) estimates that undocumented workers earn 8% less than documented foreign-born workers with the same characteristics. These estimates are very similar to those in column 5 in Table 4 (bottom panel), which belong to the model that does not include occupation fixed-effects (in line with the previous literature).<sup>18</sup>

<sup>&</sup>lt;sup>17</sup> To be clear, not all undocumented workers are employed in occupations with entry barriers. Based on our estimated conditional employment share gaps (evaluated at the overall sample mean characteristics), only 30.1% of undocumented workers are employed in occupations where they suffer a productivity loss because of lack of legal status (compared to 72.3% of employment among documented workers). That is, undocumented workers are under-represented in these occupations by a total of 42.2 percentage points (as shown in the last row of Table 3).

<sup>&</sup>lt;sup>18</sup>The estimates in Peri and Zaiour (2021) depart somewhat from the previous estimates. Using more recent data (CPS 2019-2020), they report that the average undocumented worker earns 22 log points less than the average documented foreign-born worker. This unconditional wage gap is much smaller than what one obtains using data for earlier years. Importantly, their data refers to the post-DACA period, which has been shown to have a large positive effect on the labor market outcomes of Dreamers (Pope, 2016; Amuedo-Dorantes and Antman, 2017). Conditional on individual characteristics, Peri and

#### 6.3 Heterogeneity

Clearly, the estimated lower bound is an average of the corresponding values for all occupations where undocumented workers suffer a loss in productivity. Next, we explore the heterogeneity across this set of occupations. To do this, we consider the top 20 occupations by the size of their entry barriers (as listed in Table 3) and estimate a series of models on subsamples of the data. More specifically, for each of these occupations (e.g. *Teachers & Instructors*) we consider a sample including all workers in that particular occupation together with all workers employed in occupation-specific lower bounds for the productivity penalty. The results are collected in Table 5. The estimates reveal large variation across occupations. We estimate that the largest (lower bounds) for the productivity penalty are found among *Nurses and Therapists* (32 log points), *Mail carriers and postal service workers* (28 log points) and *Managers in education and farms* (28 log points).

All in all, these estimates show that undocumented workers employed in occupations with significant entry barriers earn lower hourly wages than observationally similar documented workers. However, as our theory made clear, these wage gaps reflect several factors, including pay discrimination, productivity loss due to lack of legal status and self-selection across occupations. We estimate the undocumented productivity loss to be upwards of 12% (relative to observationally similar documented workers).

#### 6.4 Back-of-the-envelope estimation of GDP gains

Our stylized extension of the Roy model is not well suited to quantify the effects of legalization on GDP. For instance, wages are taken as given and there is no capital in the model. However, it is not difficult to incorporate the main insights from our occupation choice model into the theoretical frameworks commonly used to analyze the economic effects of immigration policies.

Specifically, our starting point is the framework used in Edwards and Ortega (2017). Their analysis of the economic contribution of undocumented workers is conducted using a standard general equilibrium model where output is produced with a constant-returns-to-scale Cobb-Douglas production function combining capital and labor, which in turn is a multi-level CES aggregate of labor inputs. In their model, the long-run effect on GDP

Zaiour (2021) estimate a residual wage gap of only 1 log point (though close to 4 log points for males), implying that practically all the raw wage gap is due to individual characteristics.

of a change in the labor input (defined as the effects once capital has adjusted to return to its initial marginal product) is proportional to the change in labor (in efficiency units). In other words, the percent change in GDP between the baseline and the legalization scenario can be written as  $g^Y = g^L$ , where  $g^L$  is the percent increase in the efficiency units of labor.

As we discussed in Section 3.6, within our framework, legalization increases the productivity of undocumented workers in two ways. Undocumented workers previously employed in occupation 2 (that is, the occupation with entry barriers) that remain in this occupation experience a productivity gain measured by  $\phi$ . This is illustrated in Figure 3 by the flat segment pertaining to the individuals with the highest values of  $\varepsilon_2$  (with measure  $Pr(U_2)$ ). The second channel of productivity improvements corresponds to the individuals that switch from occupation 1 to occupation 2 when they gain legal status (with measure Pr(MM)), illustrated by the middle segment in the figure. Occupation switchers not only use their idiosyncratic abilities more efficiently after legalization. They also enter occupations that may have higher average productivity. The overall expression for the percent change in efficiency units of labor and, hence, GDP was derived earlier (Equation 3.6) but we reproduce it below for convenience:

$$g^{Y} = u \left( \phi Pr(U_{2}) + (\mu_{2} - \mu_{1})Pr(MM) + \int_{-\infty}^{\infty} \int_{\varepsilon_{1} - (\mu_{2} - \mu_{1})}^{\phi + \varepsilon_{1} - (\mu_{2} - \mu_{1})} (\varepsilon_{2} - \varepsilon_{1}) f(\varepsilon_{1}, \varepsilon_{2}) d\varepsilon_{1} d\varepsilon_{2} \right).$$
(26)

As we explain below, some terms in the expression can be directly measured in the data (or follow from our empirical analysis). Only one term depends on distributional assumptions. Here we will specialize the joint ability distribution to the uniform case, which will allow us to quantify the percent change in GDP  $(g^Y)$ .

The first term, u, is the share of undocumented workers in the workforce (prior to legalization). According to our data, undocumented workers make up around 5% of employment (u = 0.05). In addition, the undocumented productivity penalty is upwards of 12% ( $\phi \ge 0.12$ ), the fraction of undocumented workers employed in occupations with entry barriers is about 30% (so that  $Pr(U_2) = 0.30$ ), and undocumented workers are under-represented in those occupations by 42.2 percentage points (Pr(MM) = 0.42).<sup>19</sup>

The term  $(\mu_2 - \mu_1)$  is the average productivity difference between occupations with entry barriers (for undocumented workers) and occupations without. To estimate this

<sup>&</sup>lt;sup>19</sup>See footnote 17 in regards to the estimated probabilities used here.

term we consider the subsample of documented workers in our data and estimate the average hourly wage for the two groups of occupations.<sup>20</sup> The raw averages imply that productivity (hourly wages) are 26 log-points higher in occupations with entry barriers. Conditioning on the same controls as in Equation (24), we estimate that the productivity gap falls to 13 log points. This is our preferred estimate for  $(\mu_2 - \mu_1)$ .

Last, we assume that the joint distribution for idiosyncratic ability is distributed uniformly over the unit interval  $(0 \le \varepsilon_1, \varepsilon_2 \le 0)$ . In this case, the third term within the parenthesis in Equation (26) simplifies to  $\frac{1}{6}(1 + (1 - \phi)^2(4\phi - 1))$ , as derived in Section C.2, which is solely a function of  $\phi$ .

Combining the above terms, we obtain

$$g^{Y} = 0.05 \left( 0.036 + 0.100 + 0.055 \right) = 0.96\%.$$
<sup>(27)</sup>

In words, we estimate that legalization would increase the GDP of the United States annually by 0.96% *at a minimum*. Given that the U.S. GDP is around \$21 Trillion, this amounts to a minimum increase of \$202 billion per year.

It is worth emphasizing that the contribution arising from occupation switching is quantitatively important. Failing to account for this (that is, imposing Pr(MM) = 0) would cut down the estimated lower bound for GDP growth to  $g^Y = 0.05 \times 0.036 =$ 0.18%.<sup>21</sup>

### 7 Conclusions

In policy discussions around legalization of undocumented workers, one of the most prominent and controversial issues is whether legalization entails an increase in GDP and, if so, of what magnitude. A large number of studies quantify this effect using calibrated general equilibrium models (e.g. Edwards and Ortega (2017) and Peri and Zaiour (2021)). In these analyses, the crucial parameter is the size of the labor productivity increase accompanying legalization. This parameter is typically calibrated on the

 $<sup>^{20}</sup>$ As in the previous section, we consider the top 20 occupations with the largest entry barriers versus all other occupations. Note also that we are equating the wage gap between the two groups of occupations with the gap in average productivity, which is a reasonable assumption for documented workers.

 $<sup>^{21}</sup>$ Edwards and Ortega (2017) estimated that legalization would increase U.S. GDP by 1% of GDP. They assumed that the entirety of the wage gaps between documented and undocumented workers were due to the productivity loss associated with lack of legal status, which overestimates the gains, but used a model that did not consider occupational switching.

basis of empirical estimates of the wage gaps between documented and undocumented workers, or the within-person change in the wages of undocumented workers following legalization. Implicitly, this practice equates wages and productivity (either in levels or in changes). As several authors have pointed out, undocumented workers' wages are likely to reflect employer exploitation, which probably changes discontinuously when gaining legal status, and endogenous sorting across occupations.

This paper has developed a model-based strategy to obtain a lower bound for the productivity penalty associated with lack of legal status in the context of a model where undocumented workers may be subject to pay discrimination and self-select across occupations. Our analysis has found that lack of legal status lowers the productivity of a substantial proportion of undocumented workers by *at least* 12%. This finding implies that legalizing undocumented workers entails a net gain in GDP. More specifically, we estimate that legalizing all undocumented workers would increase U.S. GDP upwards of 0.96% (about \$202 billion) per year.

There is another sense in which our estimate provides a lower bound for the productivity increase that we should expect if undocumented workers obtain legal status. Several authors have provided evidence that undocumented youth under-invest in human capital because of the uncertain returns to their educational investments, as shown by Kuka et al. (2020) and Liscow and Woolston (2018). Moreover, it has also been shown that lack of legal status reduces productivity through other channels, such as increases in stress and anxiety (Hainmueller et al., 2017; Patler and Pirtle, 2018).

We conclude by pointing out some limitations of our analysis in order to provide guidance to future researchers. Our theoretical framework is static and takes labor prices as given. The previous paragraph has already noted an important dynamic consideration that has been left out in our analysis: incentives to human capital investments. In regards to the fixed-wages restriction, we note that legalization is likely to have effects on wages. For instance, redistributing income from employers to workers may lead to increases in aggregate consumption if the latter group has a larger propensity to consume. These effects are probably negligible at the national level, but may be important in more narrowly defined labor markets. Furthermore, the occupational shifts associated with legalization may also affect the wage structure, presumably reducing wages in occupations experiencing an inflow of workers, particularly in the case of (non-traded) service occupations (González and Ortega, 2011; Cortes and Pan, 2014; Furtado and Ortega, 2020).

## References

- Abrego, Leisy J, "Legal consciousness of undocumented Latinos: Fear and stigma as barriers to claims-making for first-and 1.5-generation immigrants," Law & Society Review, 2011, 45 (2), 337–370.
- Albert, Christoph, "The Labor Market Impact of Immigration: Job Creation versus Job Competition," American Economic Journal: Macroeconomics, January 2021, 13 (1), 35–78.
- Amuedo-Dorantes, Catalina and Francisca Antman, "Schooling and labor market effects of temporary authorization: evidence from DACA," *Journal of Population Economics*, January 2017, 30 (1), 339–373.
- \_, Cynthia Bansak, and Steven Raphael, "Gender Differences in the Labor Market: Impact of IRCA," *American Economic Review*, May 2007, 97 (2), 412–416.
- Baker, Bryan C. and Nancy Rytina, "Estimates of the Unauthorized Immigrant Population Residing in the United States: January 2012," 2013. Washington, DC: Of ce of Immigration Statistics, Policy Directorate, US Department of Homeland Security. http://www.dhs.gov/sites/default/ les/publications/ois\_ill\_pe\_2012\_2.pdf.
- Bartolucci, Cristian, "Understanding the Native-Immigrant Wage Gap Using Matched Employer-Employee Data," *ILR Review*, October 2014, 67 (4), 1166–1202.
- Blair, Peter and Bobby Chung, "Occupational Licensing Reduces Racial and Gender Wage Gaps: Evidence from the Survey of Income and Program Participation," Technical Report 2017-50, HCEO working paper 2017.
- Bleakley, Hoyt and Aimee Chin, "Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants," American Economic Journal: Applied Economics, January 2010, 2 (1), 165–192.
- Borjas, George J, "Self-Selection and the Earnings of Immigrants," American Economic Review, September 1987, 77 (4), 531–53.
- Borjas, George J. and Hugh Cassidy, "The wage penalty to undocumented immigration," *Labour Economics*, 2019, *61*, 101757.
- Brown, J. David, Julie L. Hotchkiss, and Myriam Quispe-Agnoli, "Does employing undocumented workers give firms a competitive advantage?," *Journal of Regional Sci*ence, 2013, 53 (1), 158–170.
- Brucker, Herbert, Albrecht Glitz, Adrian Lerche, and Agnese Romiti, "Occupational Recognition and Immigrant Labor Market Outcomes," *Journal of Labor Economics*, 2021, 39 (2), 497–525.
- Calvo, Janet, "Professional Licensing and Teacher Certification for Non-Citizens: Federalism, Equal Protection And A State?s Socio-Economic Interests," *Columbia Journal* of Race and Law, 2017, 8 (1).
- Center for Migration Studies, "Estimates of the Unauthorized Population," 2014. Dataset based on the augmented American Community Survey datafiles hosted by IPUMS (Integrated Public Use Microdata Series), 2010 to 2014.

- Chassambouli, Andri and Giovanni Peri, "The Labor Market Effects of Reducing the Number of Illegal Immigrants," *Review of Economic Dynamics*, October 2015, 18 (4), 792–821.
- Chiswick, Barry R, "Speaking, Reading, and Earnings among Low-Skilled Immigrants," *Journal of Labor Economics*, April 1991, 9 (2), 149–170.
- Chiswick, Barry, Yew Lee, and Paul Miller, "Family matters: the role of the family in immigrants' destination language acquisition," *Journal of Population Economics*, November 2005, 18 (4), 631–647.
- Cortes, Patricia and Jessica Pan, "Foreign nurse importation and the supply of native nurses," *Journal of Health Economics*, 2014, 37 (C), 164–180.
- Devillanova, Carlo, Francesco Fasani, and Tommaso Frattini, "Employment of Undocumented Immigrants and the Prospect of Legal Status: Evidence from an Amnesty Program??," ILR Review, August 2018, 71 (4), 853–881.
- Edwards, Ryan and Francesc Ortega, "The Economic Contribution of Unauthorized Workers: An Industry Analysis," *Regional Science and Urban Economics*, 2017, 67, 119–134.
- Furtado, Delia and Francesc Ortega, "Does Immigration Improve Quality of Care in Nursing Homes?," IZA Discussion Papers 13552, Institute of Labor Economics (IZA) July 2020.
- Gleeson, Shannon and Roberto G. Gonzales, "When Do Papers Matter? An Institutional Analysis of Undocumented Life in the United States," *International Migration*, 2012, 50 (4), 1–19.
- Gonzales, Roberto G., "Learning to Be Illegal: Undocumented Youth and Shifting Legal Contexts in the Transition to Adulthood," *American Sociological Review*, 2011, 76 (4), 602–619.
- González, Libertad and Francesc Ortega, "How Do Very Open Economies Absorb Large Immigration Flows? Evidence from Spanish Regions," *Labour Economics*, 2011, 18, 57–70.
- Hainmueller, Jens, Duncan Lawrence, Linna Martén, Bernard Black, Lucila Figueroa, Michael Hotard, Tomás R Jiménez, Fernando Mendoza, Maria I Rodriguez, Jonas J Swartz et al., "Protecting unauthorized immigrant mothers improves their children?s mental health," *Science*, 2017, 357 (6355), 1041–1044.
- Hall, Matthew and Emily Greenman, "The Occupational Cost of Being Illegal in the United States: Legal Status, Job Hazards, and Compensating Differentials," *Interna*tional Migration Review, 2015, 49 (2), 406–442.
- \_ , \_ , and George Farkas, "Legal Status and Wage Disparities for Mexican Immigrants," Social Forces, 2010, 89 (2), 491–513.
- Heckman, James J and Bo E Honore, "The Empirical Content of the Roy Model," *Econometrica*, September 1990, 58 (5), 1121–1149.
- Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," *Econometrica*, 2019, 87 (5), 1439–1474.

- Hsin, Amy and Francesc Ortega, "The Effects of Deferred Action for Childhood Arrivals on the Educational Outcomes of Undocumented Students," *Demography*, Jun 2018.
- Kleiner, Morris M. and Alan B. Krueger, "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market," *Journal of Labor Economics*, 2013, 31 (S1), 173–202.
- \_ and Evgeny Vorotnikov, "Analyzing occupational licensing among the states," Journal of Regulatory Economics, October 2017, 52 (2), 132–158.
- Kossoudji, Sherrie A. and Deborah A. Cobb-Clark, "Coming out of the Shadows: Learning about Legal Status and Wages from the Legalized Population," *Journal of Labor Economics*, July 2002, 20 (3), 598–628.
- Kuka, Elira, Na'ama Shenhav, and Kevin Shih, "Do Human Capital Decisions Respond to the Returns to Education? Evidence from DACA," *American Economic Journal: Economic Policy*, February 2020, 12 (1), 293–324.
- Liscow, Zachary and William Gui Woolston, "Does Legal Status Matter for Educational Choices? Evidence from Immigrant Teenagers," American Law and Economics Review, 09 2018, 20 (2), 318–381.
- Lozano, Fernando A. and Todd A. Sorensen, "The Labor Market Value to Legal Status," IZA Discussion Papers 5492, Institute for the Study of Labor (IZA) February 2011.
- Machado, Joel, "Dealing with Undocumented Immigrants: The Welfare Effects of Amnesties and Deportations," JODE - Journal of Demographic Economics, December 2017, 83 (4), 445–492.
- Monras, Joan, Javier Vzquez-Grenno, and Ferran Elias Moreno, "Understanding the Effects of Legalizing Undocumented Immigrants," IZA Discussion Papers 10687, Institute of Labor Economics (IZA) April 2017.
- Orrenius, Pia M. and Madeline Zavodny, "The impact of E-Verify mandates on labor market outcomes," *Southern Economic Journal*, 2015, 81 (4), 947–959.
- \_ and \_, "Irregular immigration in the European Union," Working Papers 1603, Federal Reserve Bank of Dallas January 2016.
- Ortega, Francesc, Ryan Edwards, and Amy Hsin, "The Economic Effects of Providing Legal Status to DREAMers," *IZA Journal of Labor Policy*, 2019, 9:2.
- Pan, Ying, "The Impact of Legal Status on Immigrants? Earnings and Human Capital: Evidence from the IRCA 1986," *Journal of Labor Research*, 2012, 33 (2), 119–142.
- Passel, Jeffrey S. and D'Vera Cohn, "Share of Unauthorized Immigrant Workers in Production, Construction Jobs Falls Since 2007: In States, Hospitality, Manufacturing and Construction are Top Industries," March 26 2015. Washington, D.C.: Pew Research Center.
- \_ and Rebecca L. Clark, "Immigrants in New York: Their Legal Status, Incomes and Taxes," 1998. Washington, DC: Urban Institute.
- Pastor, Manuel and Justin Scoggins, "Estimating the Eligible-to-Naturalize Population," March 8 2016. Center for the Study of Immigrant Integration.

- Patler, Caitlin and Whitney Laster Pirtle, "From undocumented to lawfully present: Do changes to legal status impact psychological wellbeing among latino immigrant young adults?," *Social Science Medicine*, 2018, 199, 39 48. The role of Racism in Health Inequalities: Integrating Approaches from Across Disciplines.
- Patt, Alexander, Jens Ruhose, Simon Wiederhold, and Miguel Flores, "International Emigrant Selection on Occupational Skills," *Journal of the European Economic As*sociation, 07 2020, 19 (2), 1249–1298.
- Peri, Giovanni and Reem Zaiour, "Citizenship for Undocumented Immigrants Would Boost U.S. Economic Growth," June 2021. Center for American Progress and UC Davis Global Migration Center.
- Pope, Nolan G., "The Effects of DACAmentation: The Impact of Deferred Action for Childhood Arrivals on Unauthorized Immigrants," *Journal of Public Economics*, 2016, 143 (C), 98–114.
- Rivera-Batiz, Francisco L., "Undocumented workers in the labor market: An analysis of the earnings of legal and illegal Mexican immigrants in the United States," *Journal of Population Economics*, 1999, 12 (1), 91–116.
- Roy, A. D., "Some Thoughts On The Distribution Of Earnings," Oxford Economic Papers, 1951, 3 (2), 135–146.
- Van Hook, Jennifer, James D. Bachmeier, Donna L. Coffman, and Ofer Harel, "Can We Spin Straw Into Gold? An Evaluation of Immigrant Legal Status Imputation Approaches," *Demography*, February 2015, 52 (1), 329–354.
- Warren, Robert, "Democratizing Data about Unauthorized Residents in the United States: Estimates and Public-Use Data, 2010 to 2013," Journal on Migration and Human Security, December 2014, 2 (4), 305–328.
- \_\_\_\_ and John Robert Warren, "Unauthorized Immigration to the United States: Annual Estimates and Components of Change, by State, 1990 to 2010," International Migration Review, Summer 2013, 47 (2), 296–329.
- Weeden, Kim A., "Why Do Some Occupations Pay More than Others? Social Closure and Earnings Inequality in the United States," *American Journal of Sociology*, 2002, 108 (1).





Figure 2: Occupational choice Undocumented workers







Figure 4: Likely unauthorized individuals, by year of arrival.



Sample	All	All	All	All	Undoc
Variable	Mean	Std. Dev.	Min	Max	Mean
Year	2011	0.819	2010	2012	2011
Undocumented	0.035	0.185	0	1	1.000
Age	44.097	12.382	18	77	36.949
Female	0.454	0.498	0	1	0.302
HSD	0.066	0.249	0	1	0.425
HSG	0.581	0.493	0	1	0.408
CoGrad	0.352	0.478	0	1	0.167
Hourly wage	23.278	20.650	0	348.901	14.163
Origin South or Central America	0.066	0.249	0	1	0.724
Origin Europe	0.022	0.146	0	1	0.031
Origin Asia or Oceania	0.048	0.213	0	1	0.158
Origin Africa	0.005	0.074	0	1	0.022
Arrival by age 10	0.870	0.336	0	1	0.050
Fluent English	0.922	0.268	0	1	0.288

Table 1: Summary statistics ACS 2010-2012.

**Notes:** Pooled data for the CMS-ACS for period 2010-2012. Unweighted statistics and the number of observations is 2,763,538 and falls to 97,549 when restricted to likely undocumented individuals. The sample restricts to full-time employed individuals (with over 30 weekly work hours), older than 18 years old. HSD is an indicator for high-school dropouts, HSG is an indicator for high-school graduation (but no more education) and CoGrad is an indicator for having completed 4 years of college (or more). All US-born individuals are assumed to be fluent in English and considered to have arrived in the country before age 10. Dreamers are defined as likely undocumented who arrived to the United States by the age of 16.

		(1)	(2)	(3) Dog	(4) 1150000	(5)
Occup.	Description	Emp (th.)	Emp (th.)	Occ. Share (%)	Occ. Share (%)	0-1000 egap (%)
23	Teachers (up to secondary) and Instructors	4,013	20	4.4	0.4	-4.1
32	Nurses and Therapists	2,458	က	2.7	0.1	-2.7
04	Managers and administrators	2,854	45	3.2	0.9	-2.3
57	Secretaries	2,445	25	2.7	0.5	-2.2
08	Accountants, Auditors and Financial specialists	2,037	9	2.3	0.1	-2.1
00	CEO, Managers and administators	2,243	27	2.5	0.5	-1.9
01	Financial Managers, Human Resources	2361	42	2.6	0.8	-1.8
48	Salespersons, Advertising	$2,\!230$	35	2.5	0.7	-1.8
47	Retail Sales Clerks, Cashiers	5,711	244	6.3	4.8	-1.5
20	Social Workers, Counselors, Clergy	1,680	17	1.9	0.3	-1.5
51	Bookkepers, billing clerks, bank tellers	1,885	30	2.1	0.6	-1.5
58	Secretaries, insurance adjustors and examiners	1,627	26	1.8	0.5	-1.3
52	Customer service reps, file clerks	1,907	43	2.1	0.8	-1.3
38	Law enforcement	1,151	0	1.3	0.0	-1.3
21	Lawyers, legal assistants	1,093	က	1.2	0.1	-1.1
50	Office supervisors	1,185	15	1.3	0.3	-1.0
30	Physicians, Pharmacists, Dentists, Nutritionists	893	1	1.0	0.0	-1.0
02	Managers in education, Farmers	1,084	16	1.2	0.3	-0.9
05	Purchasing managers, Insurance adjusters	806	$\infty$	0.9	0.2	-0.7
55	Mail carriers postal service, Dispatchers	850	12	0.9	0.2	-0.7
	All Occupations	90,189	5,091	100.0	100.0	0.0
	All Occupations $egap < 0$	63, 836	1,259	70.8	24.7	-46.0
Notes: F	<sup>2</sup> ooled data for the CMS-ACS for period 2010-2012, full-	time employ	ed individuals	, age 18-66. Two-di	igit occupations bas	ed

Table 2: Gaps in occupational shares (unconditional). Top 20 occupations by size of the gap.

thousands. Columns 3 and 4 report each occupation's share in overall employment for the corresponding group. Column 5 reports the gap in employment shares in the corresponding occupation, computed as column 4 minus column 3, that is, the difference in the on 2000-2017 ACS occupational codes (recoded by IPUMS in variable OCC). Columns 1 and 2 report estimated employment in proportion of undocumented workers in an occupation and the proportion of documented workers in that same occupation (using sampling weights). The bottom row aggregates all occupations for which  $egap_o < 0$  (not just the top 20 listed in the table). Ζ

		$A_{ge}^{(1)}$	$^{(2)}_{Age}$	(3)Fem. %	$(4) \\ Fem. \%$	(5) CoG%	(6) CoG %	(7) egap %
Occup.	Description	Doc	U-Doc	Doc	U-Doc	Doc	U-Doc	U-Doc
	All Occupations	40.0	-4.6	45.0	-17.0	35.0	-21.0	
23	Teachers (up to sec.) and Instructors	40.2	-3.1	75.6	5.3	88.0	-26.7	-2.4
08	Accountants, Auditors Fin. specialists	40.0	-5.7	56.6	-17.5	79.0	2.0	-2.4
04	Managers and administrators	42.6	-5.0	41.1	-7.1	57.8	2.1	-1.7
00	CEO, Managers and administrators	42.5	-3.2	34.4	-2.3	64.3	9.6	-1.6
48	Salespersons, Advertising	40.3	-4.5	35.4	-0.4	51.7	-10.1	-1.5
47	Retail Sales Clerks, Cashiers	38.5	-4.1	47.1	1.6	23.1	-6.1	-1.5
32	Nurses and Therapists	41.4	-5.1	87.9	-14.2	60.6	20.2	-1.3
21	Lawyers, legal assistants	40.1	-4.9	57.9	12.2	76.6	-0.9	-1.2
01	Financial Managers, Human Resources	42.1	-4.8	42.6	-13.9	58.3	14.6	-1.2
20	Social Workers, Counselors, Clergy	40.6	-1.4	64.6	-24.1	74.5	-7.7	-1.1
37	Fire fighting, prevention, and inspection	41.1	-3.6	11.5	10.8	26.5	1.1	-1.0
50	Office supervisors	42.1	-5.4	62.3	-15.8	31.1	1.2	-1.0
52	Customer service reps, file clerks	38.3	-3.4	69.2	-12.9	24.5	2.6	-1.0
58	Secretaries, ins. adjustors examiners	40.6	-5.9	79.6	-8.0	20.8	1.5	-0.9
51	Bookkeepers, billing clerks, bank tellers	40.8	-6.0	84.7	-14.7	17.8	12.9	-0.8
02	Managers in education, Farmers	42.7	-4.4	40.6	-32.1	62.8	-46.9	-0.8
57	Secretaries	42.8	-8.0	95.6	-8.1	19.9	8.1	-0.8
05	Purchasing managers, Ins. adjusters	41.2	-5.6	55.6	-14.0	48.1	4.7	-0.8
55	Mail carriers postal service, Dispatchers	43.0	-7.1	39.3	-20.8	12.6	2.3	-0.7
90	Personnel, HR, Training	40.1	-3.4	62.6	-3.6	54.5	0.4	-0.7
	All Occupations with $egap < 0$							-42.2
Notes: Poole	ed data 2010-2012, full-time employed, older th	nan 18.	Two-digi	t occupati	ons $(2000-2)$	017 ACS,	IPUMS r	ecoded OCC
Column 1 reț	ports the mean age for Doc workers. Column	1 2 repo	orts the g	gap in the	mean age (	of Undoc	workers r	elative to D
workers. Simi	ilarly, columns 3-6 report the shares of female a	and coll	ege gradı	iates, by o	ccupation, a	and the co	orrespondi	ng Undoc-D
gaps. Columi	n 7 reports the gap in conditional employmen	tt share 4ucetic	s, based .	on the esti to of reside	mation of J	probit mc	dels (for t z sampling	that particul
computed the	a average probability of being employed in eac	ch part	cular occ	upation fo	r the subsa	mple of 1	Doc and U	Indoc worke
evaluating the	e expressions at the sample means to isolate the two isolates $C_{\rm or}$ which conditional equals < 0 (not just the two	he 'pure on 20 li	o' effect fi sted in th	com lack of re table).	f legal statu	s. The b	ottom row	aggregates
- monodooo	$\sim \sim $	1 1 1		· ( man of				

Table 3: Gans in occupational shares (conditional). Ton 20 occupations hy size of the can

Dep. Var. $\ln w$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Full Sample							
Undoc	-0.40***	-0.43***	-0.40***	-0.18***	-0.03***	-0.02***	-0.02***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Undoc $\times$ Barrier	-0.02* [0.010]	-0.03*** [0.009]	-0.00 [0.009]	-0.07*** [0.008]	-0.12*** [0.008]	-0.12*** [0.008]	-0.09*** [0.007]
Observations	1,847,168	1,847,168	1,847,168	1,847,168	1,847,168	1,847,168	2,109,746
FB Sample							
Undoc	-0.37***	-0.35***	-0.30***	-0.15***	-0.08***	-0.06***	-0.06***
	[0.005]	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]
Undoc $\times$ Barrier	-0.08*** [0.010]	-0.09*** [0.010]	-0.09*** [0.010]	-0.08*** [0.009]	-0.08*** [0.009]	-0.08*** [0.009]	-0.08*** [0.009]
Observations	295,628	295,628	295,628	295,628	295,628	295,628	295,628
Occ. Barriers	Top20	Top20	Top20	Top20	Top20	Top20	Top30
State FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Age and Fem	No	No	Yes	Yes	Yes	Yes	Yes
Education FE	No	No	No	Yes	Yes	Yes	Yes
Origin Fluent Child	No	No	No	No	Yes	Yes	Yes
Occupation FE	No	No	No	No	No	Yes	Yes

Table 4: Wage gaps and lower bound productivity penalty.

**Notes:** Data 2010-2012. Full-time employed, age 19-67. The dependent variable is the log hourly wage. Full sample in top panel (with 5.3% undocumented). Bottom panel restricts to foreign-born (with 29.4% undocumented). Occupational barriers based on analysis in Table 3 for the full sample. Controls *Origin Fluent Child* mean continent of origin dummies, indicator for English fluency and indicator for having arrived in the United States before age 10, respectively. Survey weights used in estimation. Year fixed-effects always included. Heteroskedasticity-robust standard errors in brackets. \*\*\* p < 0.01, \*\* p < 0.05, \*p < 0.1.

		Lower bound
		$\phi \geq -\gamma$
Occupation	Description	
	Top20 barrier occ. plus no-barrier occ.	0.12
23	Teachers (up to secondary) Instructors	0.13
08	Accountants, Auditors Fin. specialists	0.15
04	Managers and administrators	0.04
00	CEO, Managers and administrators	0.00
48	Salespersons, Advertising	0.26
47	Retail Sales Clerks, Cashiers	0.15
32	Nurses and Therapists	0.32
21	Lawyers, legal assistants	0.26
01	Financial Managers, Human Resources	-0.01
20	Social Workers, Counselors, Clergy	0.15
37	Fire fighting, prevention, and inspection	0.18
50	Office supervisors	0.08
52	Customer service reps, file clerks	0.09
58	Secretaries, ins. adjustors and examiners	0.10
51	Bookkeepers, billing clerks, bank tellers	0.04
02	Managers in education, Farmers	0.28
57	Secretaries	0.07
05	Purchasing managers, Ins. adjusters	0.05
55	Mail carriers postal service, Dispatchers	0.28
06	Personnel, HR, Training	0.06

Table 5: Occupation-specific lower bound productivity penalty

**Notes:** Data 2010-2012. Full-time employed, age 19-67. The estimated lower bounds are based on the estimation of the model in Column 6 of Table 4. The dependent variable in that model is the log hourly wage and includes all controls specified in the footnote to that table. However, the sample used varies for each of the estimates in the table: in each case, we include all no-barrier occupations and the single occupation with entry barriers considered in each case. Occupational barriers based on analysis in Table 3. \*\*\* p < 0.01, \*\* p < 0.05, \*p < 0.1.

		(1)	(2)	(3)	(4)	(5)
		Doc	Undoc	Doc	Undoc	U-Doc
Occup.	Description	$\operatorname{Emp}(\operatorname{th.})$	$\operatorname{Emp}(\operatorname{th.})$	Occ. Share $(\%)$	Occ. Share $(\%)$	$\operatorname{Gap}(\%)$
32	Nurses and Therapists	377	n	3.1	0.1	-3.0
08	Accountants, Auditors and Financial specialists	325	9	2.7	0.1	-2.5
23	Teachers (up to secondary) and Instructors	282	20	2.3	0.4	-1.9
30	Physicians, Pharmacists, Dentists, Nutritionists	233	1	1.9	0.0	-1.9
04	Managers and administrators	321	45	2.6	0.9	-1.7
01	Financial Managers, Human Resources	294	42	2.4	0.8	-1.6
00	CEO, Managers and administators	237	27	1.9	0.5	-1.4
10	Computer Analysis	501	138	4.1	2.7	-1.4
36	Nursing aides, Health aides, Dental assistants	381	91	3.1	1.8	-1.3
57	Secretaries	183	25	1.5	0.5	-1.0
22	Teachers (post-secondary)	154	14	1.3	0.3	-1.0
51	Bookkepers, billing clerks, bank tellers	191	30	1.6	0.6	-1.0
20	Social Workers, Counselors, Clergy	159	17	1.3	0.3	-1.0
58	Secretaries, insurance adjustors and examiners	169	26	1.4	0.5	-0.9
48	Salespersons, Advertising	186	35	1.5	0.7	-0.8
47	Retail Sales Clerks, Cashiers	688	244	5.6	4.8	-0.8
52	Customer service reps, file clerks	194	43	1.6	0.8	-0.7
15	Engineers and Geologists	143	22	1.2	0.4	-0.7
21	Lawyers, legal assistants	93	က	0.8	0.1	-0.7
03	Managers Medicine occ., Managers Food Lodging	179	40	1.5	0.8	-0.7
	All Occupations	12,227	5,091	100.0	100.0	0.0
	All Occupations $egap < 0$	8,211	$1,\!486$	67.2	29.2	-38.0
Notes: ]	Pooled data for the CMS-ACS for period 2010-2012, for	eign-born, fu	ull-time emple	yed individuals, ag	e 18-66. Two-digit	

occupations based on 2000-2017 ACS occupational codes (recoded by IPUMS in variable OCC). Columns 1 and 2 report estimated employment in thousands. Columns 3 and 4 report each occupation's share in overall employment for the corresponding group. Column 5 is computed as column 4 minus column 3, that is, the difference in the proportion of undocumented workers in an

occupation and the proportion of documented workers in that same occupation, computed using sampling weights.

Table 6: Gaps in occupational shares (unconditional). Foreign-born only

# Appendix

## A Simple Numerical Example

The numerical example below illustrates that, depending on distributional assumptions regarding ability, undocumented workers may be positively or negatively selected in either occupation (or in both).

**Example 1** Consider an economy with an equal number of documented and undocumented workers. The ability distribution for documented workers,  $f(\varepsilon_1, \varepsilon_2)$  is as follows. The domain takes three values  $Df = \{(a, 2), (3, b), (8, 10)\}$  with a > 2 and  $b \ge 3$ . The probability function  $f(\varepsilon_1, \varepsilon_2) = 1/3$  for the three points in the domain. The ability distribution for the 3 undocumented workers is exactly the same. Let the productivity penalty  $be \phi = 5$  and assume  $\mu_1 = \mu_2$ .

It is straightforward to verify the following observations:

- 1. Documented workers choose occupation 2 if and only if  $\varepsilon_2 \ge \varepsilon_1$ . So 2/3 of documented workers choose occupation 2 and 1/3 choose occupation 1.
- 2. Thus  $E(\varepsilon_1|D_1) = a \text{ and } E(\varepsilon_2|D_2) = \frac{10+b}{2}$ .
- Undocumented workers choose occupation 2 if and only if ε<sub>2</sub> ≥ ε<sub>1</sub> + φ. So 2/3 of documented workers choose occupation 1 and 1/3 choose occupation 2. Note that undocumented workers with (ε<sub>1</sub>, ε<sub>2</sub>) = (8, 10) are mismatched because they choose occupation 2 due to the productivity loss from lack of legal status (productivity 10 − 5 > 3).
- 4. Thus  $E(\varepsilon_1|U_1) = \frac{a+8}{2}$  and  $E(\varepsilon_2|U_2) = b$ .
- 5. Undocumented workers are positively selected (in ability) in occupation 1 if  $s_1(\phi) = \frac{8-a}{2} > 0$ , that is, if 2 < a < 8. If  $a \ge 8$  then they are negatively selected.
- 6. Undocumented workers are positively selected (in ability) in occupation 2 if  $s_2(\phi) = \frac{b-10}{2} > 0$ , that is, if b > 10. If  $8 \le b \le 10$  then they are negatively selected.

#### Β Proofs

**Proof:** Proposition 1. The proposition follows easily from the earlier observations:  $Pr(U_1) = Pr(D_1) + Pr(MM)$  and  $Pr(D_2) = Pr(U_2) + Pr(MM)$ . However, it is instructive to provide a more direct proof of the statement.

- 1. By definition, the undocumented-documented employment share gap in occupation 1 is given by  $egap_1 = Pr(U_1) - Pr(D_1)$ .
- 2. Clearly,

$$Prob(U_1) = \int_{\varepsilon_2 = -\infty}^{\varepsilon_2 = \infty} \left[ \int_{\varepsilon_1 = \varepsilon_2 - \phi + (\mu_2 - \mu_1)}^{\varepsilon_1 = \infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 \right] d\varepsilon_2$$
(B.1)  
=  $G_1(\phi)$ . (B.2)

$$G_1(\phi). \tag{B.2}$$

- 3. Naturally, because f > 0 over all its domain, as  $\phi$  increases,  $G_1(\phi)$  will increase.
- 4. Note also that  $Pr(D_1) = G_1(0)$ . It follows that  $Pr(U_1) = G_1(\phi) > G_1(0) =$  $Pr(D_1)$ , where the strict sign follows from the positive density over the mismatch region  $MM(\phi)$  for any  $\phi > 0$ . Thus undocumented workers will be overrepresented in occupation 1, that is,  $egap_1 > 0$ .
- 5. Last, the expression for  $egap_1$  in the proposition can be derived easily from Equation (7) by keeping in mind that  $Pr(U_1) = Pr(D_1) + Pr(MM)$ , as illustrated by Figure 2. Similarly, the expression for  $egap_2$  follows from Equation (8) and  $Pr(D_2) = Pr(U_2) + Pr(MM).$

Let us now turn to the undocumented-documented gaps in employment shares in occupation 2.

- 1. By definition,  $egap_2 = Pr(U_2) Pr(D_2)$ .
- 2. Define now

$$Prob(U_2) = \int_{\varepsilon_1 = -\infty}^{\varepsilon_1 = \infty} \left[ \int_{\varepsilon_2 = \varepsilon_1 + \phi - (\mu_2 - \mu_1)}^{\varepsilon_2 = \infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 \right] d\varepsilon_1$$
(B.3)  
=  $G_2(\phi).$ (B.4)

(B.4)

3. As before, because f > 0 over its domain, as  $\phi$  increases,  $G_2(\phi)$  will decrease.

4. Note also that  $Pr(D_2) = G_2(0)$ . It follows that  $Pr(U_2) = G_2(\phi) < G_2(0) = Pr(D_2)$ . Thus undocumented workers will be under-represented in occupation 2, that is,  $egap_2 < 0$ .

This completes the proof.  $\blacksquare$ 

**Proof:** Proposition 2. The first claim in the proposition is the sufficient condition for positive selection of undocumented workers in occupation 2, stated in Equation (12). To prove this claim it is helpful to proceed in steps:

- 1. Recall from Equation (B.4) in the proof of Proposition 1 that  $Pr(U_2) = G_2(\phi)$ ,  $Pr(D_2) = G_2(0)$  and function  $G_2(\phi)$  is decreasing in  $\phi$ .
- 2. Consider now the *average ability* of undocumented workers in occupation 2:

$$E(\varepsilon_2|U_2) = \int_{\substack{\varepsilon_1 = -\infty \\ \varepsilon_1 = -\infty \\ \varepsilon_1 = -\infty}}^{\varepsilon_1 = -\infty} \left[ \int_{\substack{\varepsilon_2 = \varepsilon_1 + \phi - (\mu_2 - \mu_1)}}^{\varepsilon_2 = \infty} \varepsilon_2 \frac{f(\varepsilon_1, \varepsilon_2)}{G_2(\phi)} d\varepsilon_2 \right] d\varepsilon_1$$
(B.5)

$$= \int_{\varepsilon_1=-\infty}^{\varepsilon_1=-\infty} E(\varepsilon_2|\varepsilon_1, \varepsilon_2 \ge \varepsilon_1 + \phi - (\mu_2 - \mu_1))d\varepsilon_1 \qquad (B.6)$$

$$= \int_{\varepsilon_1 = -\infty}^{\varepsilon_1 = \infty} H_2(\phi|\varepsilon_1) d\varepsilon_1, \qquad (B.7)$$

where we defined  $H_2(\phi|\varepsilon_1) = E(\varepsilon_2|\varepsilon_1, \varepsilon_2 \ge \varepsilon_1 + \phi - (\mu_2 - \mu_1))$ . In general,  $H_2(\phi|\varepsilon_1)$  is not a monotonic function of  $\phi$ .

3. Turning now to the analogous expression for *documented* workers,

$$E(\varepsilon_2|D_2) = \int_{\varepsilon_1=-\infty}^{\varepsilon_1=\infty} \left[ \int_{\varepsilon_2=\varepsilon_1-(\mu_2-\mu_1)}^{\varepsilon_2=\infty} \varepsilon_2 \frac{f(\varepsilon_1,\varepsilon_2)}{G_2(0)} d\varepsilon_2 \right] d\varepsilon_1$$
(B.8)

$$= \int_{\substack{\varepsilon_1 = -\infty \\ q \in 1^{-\infty}}}^{\varepsilon_1 = -\infty} E(\varepsilon_2 | \varepsilon_1, \varepsilon_2 \ge \varepsilon_1 - (\mu_2 - \mu_1)) d\varepsilon_1$$
(B.9)

$$= \int_{\varepsilon_1 = -\infty}^{\varepsilon_1 = \infty} H_2(0|\varepsilon_1) d\varepsilon_1.$$
 (B.10)

4. Thus, we can write the selection term (comparing the average ability of undocumented and documented workers) in occupation 2 as

$$s_2(\phi) = E(\varepsilon_2|U_2) - E(\varepsilon_2|D_2)$$

$$(B.11)$$

$$f^{\varepsilon_1 = \infty}$$

$$= \int_{\varepsilon_1 = -\infty}^{1} \left[ H_2(\phi|\varepsilon_1) - H_2(0|\varepsilon_1) \right] d\varepsilon_1.$$
 (B.12)

- 5. Clearly, the sufficient condition in Equation (12) implies that  $H_2(\phi|\varepsilon_1) \ge H_2(0|\varepsilon_1)$ (with strict sign for a positive measure of values of  $\varepsilon_1$ ). In other words,  $s_2(\phi) > 0$ .
- 6. More generally, the same condition implies the monotonicity result. If  $\phi_1 > \phi_0 > 0$ then  $s_2(\phi_1) > s_2(\phi_0)$ .

Let us turn now to the **second claim** in the proposition, the sufficient condition for negative selection of undocumented workers in occupation 1 (Equation (13)). Essentially, the proof follows the same steps used for the first claim:

- 1. As before, Equation (B.2) in the proof of Proposition 1 showed that  $Pr(U_1) = G_1(\phi)$ ,  $Pr(D_1) = G_1(0)$  and function  $G_1(\phi)$  is increasing in  $\phi$ .
- 2. Consider the *average ability* of undocumented workers in occupation 1:

$$E(\varepsilon_1|U_1) = \int_{\substack{\varepsilon_2 = -\infty \\ \varepsilon_2 = -\infty \\ \varepsilon_2 = -\infty}}^{\varepsilon_2 = -\infty} \left[ \int_{\varepsilon_1 = \varepsilon_2 - \phi + (\mu_2 - \mu_1)}^{\varepsilon_1 = \infty} \varepsilon_1 \frac{f(\varepsilon_1, \varepsilon_2)}{G_1(\phi)} d\varepsilon_1 \right] d\varepsilon_2$$
(B.13)

$$= \int_{\substack{\varepsilon_2 = -\infty \\ \varepsilon_2 = -\infty \\ \varepsilon_2 = -\infty}}^{\varepsilon_2 = -\infty} E(\varepsilon_1 | \varepsilon_2, \varepsilon_1 > \varepsilon_2 - \phi + (\mu_2 - \mu_1)) d\varepsilon_2 \qquad (B.14)$$

$$= \int_{\varepsilon_2 = -\infty}^{\varepsilon_2 = -\infty} H_1(\phi|\varepsilon_2) d\varepsilon_2, \qquad (B.15)$$

where the last step defines  $H_1(\phi|\varepsilon_2)$ . As was the case before,  $H_1(\phi|\varepsilon_2)$  is not necessarily a monotonic function of  $\phi$ .

3. As before, we can write the selection term as

$$s_1(\phi) = E(\varepsilon_1|U_1) - E(\varepsilon_1|D_1)$$
(B.16)

$$= \int_{\varepsilon_2 = -\infty}^{\varepsilon_2 = -\infty} \left[ H_1(\phi|\varepsilon_2) - H_1(0|\varepsilon_2) \right] d\varepsilon_2.$$
 (B.17)

- 4. Clearly, the sufficient condition in Equation (13) implies that  $H_1(\phi|\varepsilon_2) < H_1(0|\varepsilon_2)$ (with strict sign for a positive measure of values of  $\varepsilon_2$ ). In other words,  $s_1(\phi) < 0$ , as we wanted to prove.
- 5. As before, the same condition implies the monotonicity result. If  $\phi_1 > \phi_0 > 0$  then  $s_1(\phi_1) < s_1(\phi_0)$ .

This completes the proof of Proposition 2.  $\blacksquare$ 

**Proof:** Proposition 3. The first claim specializes the joint distribution for idiosyncratic productivity (ability) to the uniform distribution. Namely, let z > 0 and assume  $f(\varepsilon_1, \varepsilon_2) = 1/z^2$  for  $0 \le \varepsilon_1, \varepsilon_2 \le z$  and zero otherwise. Rather than check the sufficient condition stated in the first claim of the proposition, it is simpler to directly compute the corresponding conditional expectations. For simplicity, we assume  $\mu_1 = \mu_2$  in the special case with uniform distribution.

We begin with the claim regarding self-selection in occupation 2. Our strategy will be to compute  $E(\varepsilon_2|U_2)$  and directly compare it to  $E(\varepsilon_2|D_2)$ . It is helpful to proceed in steps.

1. Consider first  $E(\varepsilon_2|U_2)$  for the uniform distribution:

$$E(\varepsilon_2|U_2) = \int_{\varepsilon_1=0}^{\varepsilon_1=z-\phi} \left[ \int_{\varepsilon_2=\varepsilon_1+\phi}^{\varepsilon_2=z} \varepsilon_2 \frac{f(\varepsilon_1,\varepsilon_2)}{G_2(\phi)} d\varepsilon_2 \right] d\varepsilon_1$$
(B.18)

$$= \frac{1}{G_2(\phi)} \frac{1}{z^2} \int_{\varepsilon_1=0}^{\varepsilon_1=z-\phi} \left[ \int_{\varepsilon_2=\varepsilon_1+\phi}^{\varepsilon_2=z} \varepsilon_2 d\varepsilon_2 \right] d\varepsilon_1$$
(B.19)

where  $G_2(\phi) = Prob(U_2)$ .

2. Computing the integrals delivers

$$E(\varepsilon_2|U_2) = \frac{2z^3 - 3\phi z^2 + \phi^3}{3(z-\phi)^2}.$$
 (B.20)

3. Next, recall that  $E(\varepsilon_2|D_2) = E(\varepsilon_2|U_2)$  for  $\phi = 0$ . Hence, the expression in the previous point simplifies to

$$E(\varepsilon_2|D_2) = \frac{2z}{3}.$$
 (B.21)

4. Last, it is straightforward to check that  $E(\varepsilon_2|U_2) > E(\varepsilon_2|D_2)$  if and only if  $(z - \phi)^2 > 0$ , which is obviously satisfied. This proves that undocumented workers are unambiguously positively selected in occupation 2 in terms of ability  $(s_2(\phi) > 0)$  when ability is distributed uniformly.

We now turn to self-selection in occupation 1. Proving this claim simply requires adapting the argument given for self-selection in occupation 2. In this case we shall compute  $E(\varepsilon_1|U_1)$  and directly compare it to  $E(\varepsilon_1|D_1)$ . As before, it helps to proceed in steps. 1. Consider first  $E(\varepsilon_1|U_1)$  for the uniform distribution:

$$E(\varepsilon_1|U_1) = \frac{1}{G_1(\phi)} \left( \int_{\varepsilon_1=0}^{\varepsilon_1=z-\phi} \int_{\varepsilon_2=0}^{\varepsilon_2=\varepsilon_1+\phi} \frac{\varepsilon_1}{z^2} d\varepsilon_2 d\varepsilon_1 + \int_{\varepsilon_1=z-\phi}^{\varepsilon_1=z} \int_{\varepsilon_2=0}^{\varepsilon_2=z} \frac{\varepsilon_1}{z^2} d\varepsilon_2 d\varepsilon_1 \right)$$

where  $G_1(\phi) = Pr(U_1)$ .

2. Computing the integrals delivers

$$E(\varepsilon_1|U_1) = \frac{1}{3} \frac{3z^3 - (z-\phi)^3}{z^2 - \phi^2 + 2z\phi}$$
(B.22)

$$E(\varepsilon_1|D_1) = \frac{2}{3}z, \qquad (B.23)$$

where the second expression relies on  $E(\varepsilon_1|D_1) = E(\varepsilon_1|U_1)$  for  $\phi = 0$ .

3. Last, it is straightforward to check that  $s_1(\phi) = E(\varepsilon_1|U_1) - E(\varepsilon_1|D_1) < 0$  if and only if  $\phi^2 < z^2 + \phi z$ , which always holds given that  $0 < \phi \le z$ .

Let us turn now to the second claim in the proposition. Namely, assume that  $(\varepsilon_1, \varepsilon_2)$  are distributed as a bivariate normal with means  $(m_1, m_2)$ , standard deviations  $(\sigma_1, \sigma_2)$  and correlation coefficient  $\rho$ . Thus, we are not imposing independence. To prove this claim it is helpful to use the sufficient conditions stated in Proposition 2. To lighten the equations it is helpful to define  $\gamma = \phi - (\mu_2 - \mu_1)$ .

Let us first focus in the selection claim concerning occupation 2, which entails the comparison between  $E(\varepsilon_2|U_2)$  and  $E(\varepsilon_2|D_2)$ .

1. By virtue of Equation (B.6),

$$E(\varepsilon_2|U_2) = \int_{\varepsilon_1=-\infty}^{\varepsilon_1=\infty} E(\varepsilon_2|\varepsilon_1, \varepsilon_2 \ge \varepsilon_1 + \gamma) d\varepsilon_1.$$
 (B.24)

- 2. It is well known that the conditional distributions of a bivariate normal are univariate normals. In particular,  $\varepsilon_2 | \varepsilon_1$  is distributed  $N(\hat{\mu}_2, \hat{\sigma}_2) = N(m_2 - \sigma_2 \rho \frac{\varepsilon_1 - m_1}{\sigma_1}), (1 - \rho^2)\sigma_2^2).$
- 3. For univariate normal variables there is a closed-form solution for censored expec-

tations. Specifically,

$$E(\varepsilon_2|\varepsilon_1, \varepsilon_2 \ge \varepsilon_1 + \gamma) = \frac{\phi\left(\frac{\varepsilon_1 + \gamma - \hat{\mu}_2}{\hat{\sigma}_2}\right)}{1 - \Phi\left(\frac{\varepsilon_1 + \gamma - \hat{\mu}_2}{\hat{\sigma}_2}\right)} = R\left(\frac{\varepsilon_1 + \gamma - \hat{\mu}_2}{\hat{\sigma}_2}\right),$$
(B.25)

where  $\phi$  (abusing notation) and  $\Phi$  denote the density and cdf of the standard normal distribution.

4. Expression R(x) is the inverse Mills ratio, well known to be a (strictly) increasing function for the standard normal distribution (Heckman and Honore, 1990). Since  $\gamma = \phi - (\mu_2 - \mu_1)$ , function R will also be increasing in productivity loss parameter  $\phi$ . Hence, sufficient condition Equation (12) is satisfied.

Let us now turn to the selection claim for occupation 1, which entails the comparison between  $E(\varepsilon_1|U_1)$  and  $E(\varepsilon_1|D_1)$ .

- 1. As noted above, conditional distribution  $\varepsilon_1 | \varepsilon_2$  is distributed  $N(\hat{\mu_1}, \hat{\sigma_1}) = N(m_1 \sigma_1 \rho \frac{\varepsilon_2 m_2}{\sigma_2}), (1 \rho^2) \sigma_1^2).$
- 2. Using the expression for the censored expectation of a univariate normal variable, we obtain

$$E(\varepsilon_1|\varepsilon_2,\varepsilon_1 > \varepsilon_2 - \gamma) = \frac{\phi\left(\frac{\varepsilon_2 - \gamma - \hat{\mu}_1}{\hat{\sigma}_1}\right)}{1 - \Phi\left(\frac{\varepsilon_2 - \gamma - \hat{\mu}_1}{\hat{\sigma}_1}\right)} = R\left(\frac{\varepsilon_2 - \gamma - \hat{\mu}_1}{\hat{\sigma}_1}\right),$$
(B.26)

where (still abusing notation)  $\phi$  and  $\Phi$  denote the density and cdf of the standard normal distribution.

3. As argued above, the inverse Mills ratio (R(x)) is a (strictly) increasing function for the standard normal distribution. However, because  $\gamma = \phi - (\mu_2 - \mu_1)$  and  $\gamma$  is preceded by a negative sign as an argument of R(x), the previous conditional expectation will be a decreasing function of  $\phi$ . Hence, sufficient condition Equation (13) is satisfied. **Proof:** Proposition 5. The proof of the claim relies on three simple observations:

1. Because of negative selection in occupation 1, we can bound  $wgap_1$  above by

$$wgap_1 = s_1 - \tau < -\tau. \tag{B.27}$$

2. Likewise, because of positive selection in occupation 2, we can bound  $wgap_2$  from below by

$$wgap_2 = s_2 - \tau - \phi > -\phi - \tau.$$
 (B.28)

3. Combining the previous expressions,

$$wgap_2 - wgap_1 < -\tau - (-\phi - \tau) \tag{B.29}$$

$$wgap_2 - wgap_1 < \phi. \tag{B.30}$$

## C Examples with uniform ability distribution

Assume the joint ability distribution is uniform. Namely, let z > 0 and assume  $f(\varepsilon_1, \varepsilon_2) = 1/z^2$  for  $0 \le \varepsilon_1, \varepsilon_2 \le z$  and zero otherwise. Further, we shall impose z = 1.

#### C.1 Productivity lower bound

It follows from equations Equation (B.20) and Equation (B.21) that

$$s_2(\phi) = E(\varepsilon_2|U_2) - E(\varepsilon_2|D_2) = \frac{1}{3} \frac{2z^3 - 3\phi z^2 + \phi^3}{(z - \phi)^2} - \frac{2z}{3}.$$
 (C.1)

Analogously, equations Equation (B.22) and Equation (B.23) imply that

$$s_1(\phi) = E(\varepsilon_1|U_1) - E(\varepsilon_1|D_1) = \frac{1}{3}\frac{3z^3 - (z-\phi)^3}{z^2 - \phi^2 + 2z\phi} - \frac{2}{3}z.$$
 (C.2)

It is easy to see that specializing to z = 1, the condition for an informative lower bound for  $\phi$  reduces to

$$\phi > s_2(\phi) - s_1(\phi)$$
 (C.3)

$$\phi > \frac{1}{3} \left( \frac{2 - 3\phi + \phi^3}{(1 - \phi)^2} - \frac{3 - (1 - \phi)^3}{1 - \phi^2 + 2\phi} \right)$$
 (C.4)

It is easy to check (numerically) that this condition holds over the interval  $\phi \in [0, 2]$ . This interval contains the range of plausible values for documented-undocumented wages. Recall that  $\phi > 0$  is a wedge between the *log* wages of documented and undocumented workers in occupation 2 ( $\omega^{2D} - \omega^{2U}$ ). Thus it is a wedge in the corresponding relative wages ( $w_2^D/w_2^U$ ). Clearly,  $\phi \in [0, 2]$  if and only if the documented-undocumented wage ratio  $\frac{w_2^D}{w_2^U} \in [1, 5.4]$ . It is implausible to expect the wages of documented workers to be more than 5 times the wages of undocumented workers within the same occupations (and after controlling for educational attainment and other sociodemographic characteristics).

### C.2 GDP gains from legalization

The goal here is to derive a closed-form solution for the GDP gains from legalization of undocumented workers. As above, we assume that the joint distribution of ability is uniform (with parameter z = 1). Furthermore, we also assume  $\mu_2 = \mu_1$ . In this case, Equation 3.6 can be written as

$$g^{Y}(\phi) = u\left(\phi Pr(U_2) + A + B\right) \tag{C.5}$$

where

$$A = \int_{\varepsilon_1=0}^{\varepsilon_1=1-\phi} \int_{\varepsilon_2=\varepsilon_1}^{\varepsilon_2=\phi+\varepsilon_1} (\varepsilon_2-\varepsilon_1) d\varepsilon_1 d\varepsilon_2$$
(C.6)

$$B = \int_{\varepsilon_1=1-\phi}^{\varepsilon_1=1} \int_{\varepsilon_2=\varepsilon_1}^{\varepsilon_2=1} (\varepsilon_2-\varepsilon_1) d\varepsilon_1 d\varepsilon_2.$$
(C.7)

Computing the integrals delivers

$$A = \phi(1-\phi)(1-\frac{\phi}{2})$$
 (C.8)

$$B = \frac{\phi}{2} - \frac{1}{2} \left( 1 - (1 - \phi)^2 \right) + \frac{1}{6} \left( 1 - (1 - \phi)^3 \right).$$
(C.9)

Adding them up we obtain

$$A + B = \frac{1}{6} + \frac{1}{6}(1 - \phi)^2(4\phi - 1).$$
 (C.10)

Using now that  $Pr(U_2) = \frac{(1-\phi)^2}{2}$ , the GDP gains from legalization simplify to

$$\Delta GDP(\phi) = u\left(\frac{1}{6} + \frac{1}{6}(1-\phi)^2(7\phi-1)\right), \qquad (C.11)$$

which can easily be checked to be a positive number for all  $0 \le \phi \le 1$ .