Self-Employment Rates of Immigrants and Natives: Occupations and Skills

Nick Manuel
nmanuel@uoguelph.ca

October 28, 2019

Abstract

It has become well established that immigrants to the United States are more likely to be self-employed than native born Americans. However, the reasons behind this difference in self-employment rates remains an open question, with researchers to date finding different and often contradictory results. In this paper, I make two main contributions to the literature. I provide the first evidence that approximately half of the gap in self-employment rates between immigrants and natives occurs within occupations. Furthermore, I find evidence that although they do not work in occupations that are more highly valued in self-employment, immigrants are more likely than natives to choose the sector (self or wage employment) in which their earnings potential is estimated to be higher. The fact that immigrants place a higher value on earnings potential when selecting self or wage employment explains approximately one-third of the gap in self-employment rates between immigrants and natives.


Keywords: Immigration, Returns to Skills, Self-Employment, Immigrant Self-Employment.
1 Introduction

Researchers have long noted that immigrants to developed countries are more likely to be self-employed than natives, a finding that dates back as early as Borjas (1986), and has continued to be found using more recent data (Li 2001, Hunt 2011, Fairlie and Lofstrom 2015, Kerr and Kerr 2016, Green et al. 2016). It is worthwhile noting that this gap in self-employment rates occurs despite evidence that immigrants face several disadvantages in becoming self-employed. These disadvantages include greater difficulty accessing credit (Blanchflower et al. 2003, Alden and Hammarstedt 2016), and a low likelihood of inheriting a family firm\footnote{While children of immigrant entrepreneurs frequently enter the same line of work as their parents (Andersson and Hammarstedt 2010), they are not considered to be immigrants in the majority of the aforementioned papers since most of them were either born in the United States or arrived as children. Nor will they be included in the immigrant sample since I exclude immigrants who arrived prior to the age of 18. The odds of an immigrant having inherited a family business are expected to be negligible.}

The academic debate surrounding the source of this immigrant-native gap in self-employment rates has primarily been framed in terms of the ”push” and ”pull” theories of self-employment\footnote{The terms ”necessity” and ”opportunity” are often used in place of push and pull.} (Clark and Drinkwater 2000, Abada et al. 2014, Fisher and Lewin 2018). According to the push theory of self-employment, immigrants are forced to become self-employed since they face poor wage employment opportunities (Volery 2007, Adada et al. 2014, Fisher and Lewin 2018, Green et al. 2018). However, proponents of the pull theory argue that immigrants have advantages in self-employment, as opposed to disadvantages in wage employment (Yuengert 1995, Hammarstedt and Shukur 2009, Vandor and Franke 2016, Vinogradov and Jorgensen 2017). However the empirical research that has attempted to separate between these two potential explanations has produced mixed results, with each of the aforementioned papers findings results supporting the theory in question. Indeed, it has been argued by Dawson and Henley (2012) that the distinction between the push and pull theories of self-employment is arbitrary, since in either case immigrants choose their sector so as to maximize their earnings.

It has also been postulated that immigrants may be more likely to be self-employed than natives as a result of stronger preferences for self-employment, either due to strong self-employment traditions in their country of origin (Yuengert 1995), or because immigrants have a higher tolerance for risk (Andersson and Hammarstedt 2010, Batista and Umbljijs 2014). However, the evidence for these theories has also been mixed. Fairlie and Meyer (1996) find that immigrants who come from
countries with higher rates of self-employment are no more likely to be self-employed after migrating to the U.S. than immigrants from countries with lower self-employment rates, however Hammarstedt and Shukur (2009) find the opposite result among immigrants to Sweden. Although there have been no empirical studies that attempt to explain this gap in self-employment rates using risk preferences, the literature comparing the risk tolerances of immigrants and natives has itself produced mixed results. Consistent with expectations, Jaeger et al. (2010) and Bauernschuster et al. (2014) find that individuals who migrate across countries or across cultural barriers have higher tolerances for risk. However, in contrast, Bonin et al. (2009) find that immigrants to Germany are more risk averse than native born Germans.

In this paper, I contribute to the literature by exploring a new potential explanation for the immigrant-native gap in self-employment rates. Taking inspiration from the burgeoning literature on occupational skills (Autor et al. 2003, Acemoglu and Autor 2011), I estimate whether differences in the skill composition between immigrants and natives can explain the higher self-employment rates among immigrants. Furthermore, I estimate whether immigrants and natives are equally likely to choose self-employment on the basis of how their skills are valued in self-employment relative to wage employment. Although many papers have explored the role of skills in explaining labour market outcomes between immigrants and natives (Fortin et al. 2016, Imai et al. 2018), this paper is the first to do so in the context of immigrant self-employment.

The best known work on the topic of skills and self-employment is perhaps Lazear (2004). In this seminal theory he postulates that entrepreneurs are “generalists” who require a certain base level of many different skills that are required in the operation of their business, while in contrast employees are “specialists”, who are highly skilled in a small number of areas. More recent work on skills and self-employment has found that entrepreneurs receive higher returns to their education than employees (Van Praag et al. 2013), and experience higher returns to capital and skills (Falco and Haywood 2016). Furthermore, Hartog et al. (2010) find that self-employed individuals receive higher returns to their social, technical and mathematical skills than employees, while employees receive higher returns to their verbal and clerical skills.

---

3It should also be noted that these two papers use different methodologies. The difference in their results is not necessarily due to differences between immigrants in the U.S. and Sweden.
The results from this literature suggest that skills may be valued differently for the self-employed in comparison to the wage employed. This paper provides the first estimates of the value of occupational skills in each sector, where occupational skills are measured using data from O*NET. Since individuals may choose self or wage employment on the basis of the earnings potential of their skills in each sector, I correct for potential non-random selection into self-employment using the classic sample selection approach of Heckman (1977). Using these estimates of skill values for each sector, I am able to estimate whether the gap in self-employment rates between immigrants and natives can be explained by differences in their skill compositions or differences in the degree of emphasis that they place on their skills valued when selecting a sector.

It is found that the higher observed self-employment rates among immigrants are not entirely the result of immigrants being more likely to enter occupations that are traditionally associated with self-employment. Rather, immigrants are on average more likely to be self-employed than natives who work in the same occupation. Notable occupations for which immigrants are more likely to be self-employed than natives are taxi drivers and chauffeurs, and food service managers. These findings suggest that different patterns of occupational choice between immigrants and natives does not entirely explain the gap in self-employment rates. Instead, it appears that a significant portion of this gap occurs because, for certain occupations, immigrants are more likely to become self-employed than their native counterparts.

Furthermore, the results suggest that only a small portion of the gap in self-employment rates can be explained by differences in skill composition between immigrants and natives. However, immigrants are more likely than natives to choose self-employment when their skills carry a relatively low value in wage employment, a response which explains approximately one-third of the gap in self-employment rates. These results indicate that the higher self-employment rates among immigrants may be the result of immigrants placing a higher value than natives on potential earnings opportunities when selecting a sector.

However, the results also contradict some of the most commonly cited explanations for the gap.

451.64% of immigrant taxi drivers report being self-employed, compared to 24.94% of native taxi driver. To check whether the higher self-employment rate among immigrants taxi drivers may be explained by a higher uptake of Uber among immigrants, I estimate the same gap using earlier waves of the ACS. While the gap has grown (consistent with an Uber explanation), a notable gap was also present being the widespread adoption of Uber.
in self-employment rates between immigrants and natives. I find that immigrants who have skills that are highly valued in wage employment are more likely to choose wage employment than similar natives, contrary to the expectations of the ”push” hypothesis. Furthermore, since the higher self-employment rates among immigrants is largely confined to those immigrants with poor opportunities in wage employment, given their occupation, I believe that this indicates that immigrants do not have a stronger preference for self-employment itself. With stronger preferences for self-employment, we would anticipate that the gap in self-employment rates would persist across the skill distribution.

2 Data and Descriptive Statistics

The data consists of the American Community Survey 2016 1-year Person Files. My primary population of interest consists of individuals who earn a living working full-time in either self-employment or wage employment. Therefore, I restrict the sample to eliminate individuals who report working fewer than 20 hours per week on average. I also exclude individuals with an average hourly income of less than $7.25\textsuperscript{5} or more than $100\textsuperscript{6} Furthermore, I exclude individuals without a recorded occupation, as well as immigrants who arrived in the United States prior to the age of 18, since they are expected to possess characteristics that similar to those of natives relative to the characteristics of other immigrants. After imposing these restrictions, I find a substantial gap in self-employment rates between immigrants and natives that is consistent with earlier work in the literature, with 7.5% of natives being self-employed as opposed to 10.3% of immigrants (see Figure 1).

While there exists a large literature examining self-employment rates across demographic groups such as age (Blanchflower (2000), Zissimopoulos and Karoly (2007)), education (Blanchflower (2000), Parker (2009)), and in particular ethnicity (Fairlie and Meyer (1996,2000), Bogan and Darity (2008)), it is worthwhile noting that the immigrant-native differences in self-employment rates are largely separate from differences in these characteristics. While I observe striking differences in self-employment rates across these characteristics (as shown in Figures 2,3,4 & 5) it is equally striking that within each of the examined demographic groups, immigrants are more likely to be self-employment than natives (with the exception of individuals with graduate or professional degrees). I control for a rich

\textsuperscript{5}The U.S. federal minimum wage

\textsuperscript{6}I exclude very low and very high earners in order to ensure that the results are not driven by outliers. This is especially important since self-employment earnings distribution contains heavier tails, with both a large low earnings informal sector (Rajiman (2001)) as well as an over-representation of very high earners.
array of demographic characteristics in a probit model of the form:

\[ \Phi = \alpha_1 + \beta_{x} \cdot X_i + \beta_{imm} \cdot Imm_i + \epsilon_i \]

Where \( \Phi \) is a latent variable, with individuals choosing self-employment if \( \Phi > 0 \), or wage employment if \( \Phi < 0 \). \( X_i \) is a rich set of demographic controls. Including these controls for the probability of being self-employed fails to explain a large portion of the raw gap in self-employment rates between immigrants and natives. The estimated margins from this model, shown in Figure 6, yield predicted self-employment rates of 7.5% and 9.6% for natives and immigrants respectively.

What is perhaps even more striking is that even though there have been well documented differences in the occupational distribution of immigrants relative to natives, differences in industry and occupation are only able to explain approximately half of the gap in self-employment rates between immigrants and natives. As shown in Figure 7, adding controls for two digit NAICS industry codes and detailed Standard Occupational Classification (SOC) codes to the model explains slightly more than half of the immigrant-native gap in self-employment rates that remains after accounting for demographics. Even within the same industry and occupation, my results show that approximately 7.5% of natives are expected to be self-employed, in comparison to 8.4% of immigrants.

These results suggest that the common idea that immigrants to OECD countries are more likely to be self-employed because they disproportionately work in occupations that are related to businesses such as corner stores, restaurants, nail salons, or motels, (Patel and Vella (2013), Kerr and Mandorff (2015)) provides only a partial picture of immigrant self-employment in the United States. While I do find that immigrants are more likely to be working in occupations that are associated with high self-employment rates, they are also more likely to be self-employed than otherwise comparable natives working in the same occupation. Notable examples of occupations in which immigrants are more likely to be self-employed than natives are presented in Figure 8.

---

7The full set of characteristics, along with differences in these characteristics between immigrants and natives, can be found in Table 1.
8In addition to controlling for SOC codes, I also estimate a model that controls for the OCC codes that are included in the ACS data. These two models generate similar results.
9Notable examples include Miscellaneous Personal Appearance Workers (57.4% immigrants), Tailors, Dressmakers and Sewers (44.6% immigrants), Taxi Drivers and Chauffeurs (36.5% immigrants), Laundry and Dry Cleaning Workers (30.3% immigrants) and Chefs and Head Cooks (28.2% immigrants).
The focus of the remainder of this paper is to attempt to explain why it is the case that immigrants are more likely to be self-employed within occupations.

Table 1: Demographic Characteristics of Natives and Immigrants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>% of Natives</th>
<th>% of Immigrants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>49%</td>
<td>45.7%</td>
</tr>
<tr>
<td>Married</td>
<td>53.5%</td>
<td>70.1%</td>
</tr>
<tr>
<td>Less Than High School</td>
<td>9.3%</td>
<td>25.1%</td>
</tr>
<tr>
<td>High School</td>
<td>20.7%</td>
<td>17.1%</td>
</tr>
<tr>
<td>Below Bachelors</td>
<td>34.4%</td>
<td>18.9%</td>
</tr>
<tr>
<td>Bachelors</td>
<td>22.3%</td>
<td>20.3%</td>
</tr>
<tr>
<td>Above Bachelors</td>
<td>13.3%</td>
<td>18.7%</td>
</tr>
<tr>
<td>Indigenous</td>
<td>3.7%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Latin American</td>
<td>27%</td>
<td>48.5%</td>
</tr>
<tr>
<td>Middle Eastern</td>
<td>0.5%</td>
<td>3.9%</td>
</tr>
<tr>
<td>European</td>
<td>52.9%</td>
<td>6.4%</td>
</tr>
<tr>
<td>African</td>
<td>8.7%</td>
<td>4.4%</td>
</tr>
<tr>
<td>East Asian</td>
<td>1.8%</td>
<td>20.4%</td>
</tr>
<tr>
<td>South Asian</td>
<td>0.2%</td>
<td>9%</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>5.1%</td>
<td>7.3%</td>
</tr>
<tr>
<td>Homeowner</td>
<td>62.7%</td>
<td>50.9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Native Mean</th>
<th>Immigrant Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>43.974</td>
<td>47.08</td>
</tr>
</tbody>
</table>

Figure 1: Raw Gap in Self-Employment Rates
3 Theory

4 Model

Lucas (1978) developed an early model of self-employment entry, by allowing for potential self-employment earnings to vary based on an individual’s managerial ability. This model predicts that individuals with higher levels of managerial ability are more likely to be self-employed and to manage
larger firms conditional on being self-employed. Evans and Jovanovich (1989) expand upon the work of Lucas by introducing credit constraints, and by allowing wage sector earnings to vary based on an individual’s characteristics. This model predicts, similar to Lucas (1978), that on average individuals with higher levels of entrepreneurial ability will be more likely to be self-employed, however rather than a strict cutoff level of managerial ability at which individuals choose to become self-employed, variations in access to credit and potential wage sector earnings allow for differences in
self-employment entry conditional on a given level of entrepreneurial ability. More recent work by Hamilton et al. (2019) and Humphries (2016), introduce non-pecuniary benefits into the decision of whether to become self-employed, in addition to earnings potential. The latter work by Humphries (2019) is a dynamic model, that allows for differences in self-employment rates across the lifecycle.

In this paper, I develop a model that is based upon the work of the above authors, and introduce
a few modifications. First, for individuals who choose self-employment, I introduce hard capital requirements that vary by occupation. Second, since I am interested in gap in self-employment rates within occupation, I explicitly allow for individuals to enter a given occupation in either the self-employment sector or the wage employment sector. Third, assign a set of skills to each occupation, and allow the returns to these skills to differ between immigrants and natives.

4.1 Model Outline

Individual i chooses from a set of occupations $d = (1, ..., D)$ and from sectors $j = (se, we)$ where $we$ denotes wage employment and $se$ denotes self-employment. If the individual chooses occupation $d$ and chooses to be wage employed, their income will be:

$$Y^i_{dwe} = W(X_i, \eta_d) \epsilon^i_{dwe}$$

Where $X_i$ denotes a set of individual characteristics, $\eta_d$ is a vector of A skills that are associated with occupation $d$, $\eta_d = (\eta_{1d}, \eta_{2d}, ..., \eta_{Ad})$. $\epsilon^i_{dwe}$ represents an error term of mean 1.

If they are self-employed, their earnings are given by:
\[ Y_{dse}^i = \Theta(X_i, \eta_d) \hat{k}_d^{\alpha_d} \epsilon_{dse}^i - r(X_i) \hat{k}_d \]

Where \( \hat{k}_d \) denotes the amount of capital that the individual requires in occupation \( d \). \( \alpha_d \) represents the occupation specific return on capital, and \( r(X_i) \) denotes the rental rate of capital, which may vary on the basis of individual characteristics. Let \( k_d \) denote the capital requirement of occupation \( d \) for self-employment.

\[
\hat{k}_d = k_d \text{ if } \lambda_i \ast \beta \ast q_i > k_d \\
0 \text{ otherwise}
\]

Where \( q_i \) denotes the characteristics of individual \( i \)'s assets. \( \beta \) is a coefficient that assigns a value to these characteristics. An example of \( q_i \) and \( \beta \) would be \( q_i \) represents a set of characteristics of an individual’s home, with \( \beta \) reflecting local housing prices. \( \lambda_i \) indicates the degree to which individuals are credit constrained.

When employed in occupation \( d \) and sector \( j \), individual \( i \) receives a utility of:

\[
u_{i}^{d,j} = Y_{i}^{d,j} + v_{i}^{d,j}
\]

Where \( Y_{i}^{d,j} \) is their earnings potential in pair \( d, j \), and \( v_{i}^{d,j} \) indicates their non-pecuniary benefit of working in occupation \( d \) and sector \( j \).

Prior to making an occupation/sector choice, I assume that an individual knows their non-pecuniary benefit \( v_{i}^{d,j} \) from pair \( d, j \). However, they do not know the disturbance \( \epsilon_{d,j}^i \) in their earnings function.

The individual chooses the occupation/sector pair that maximizes their expected utility according to

\[
\arg\max_{d,j} \ E_i[u_{d,j}^i]
\]

Where
\[
E_i[u^d_{ij}] = E_i[\rho Y^d_{ij} + v^d_{ij}] = E_i[\rho Y^d_{ij}] + E_i[v^d_{ij}] = \rho E_i[Y^d_{ij}] + v^d_{ij}
\]

and

\[
E_i[Y^d_{ij}] = W(X_i, \eta_d) \text{ if } j = we
\]

\[
\quad = \theta(X_i, \eta_d)k^d - r(M_i)k_d \text{ if } j = se
\]

From the point of view of the econometrician, the probability that an individual with characteristics \(X_i\) chooses occupation/sector pair \(d, j\) is given by:

\[
\text{Prob}(\rho E_i[Y^d_{ij}] + v^d_{ij} > \rho E_i[Y^p_{ir}] + v^p_{ir} \quad \forall \quad p, r \text{ s.t. } p \neq d \text{ and/or } r \neq j)
\]

That is the probability that \(d, j\) is chosen is equal to the probability that \(d, j\) yields a higher utility than all other possible occupation/sector pairs.

The expected probability that an individual chooses pair \(dq, se\) (where \(q\) represents an arbitrary occupation) is given by:

\[
Pr(j = se & d = dq) = \int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{dqse}(X - \rho E_i[Y^{dqse}_i] * I(z \geq \frac{k_d}{\lambda^{d\beta}})) \prod_{p \neq d} F_{psc}(X - \rho E_i[Y^{psc}_i] * I(z \geq \frac{k_p}{\lambda^{ps\beta}})) \prod_{q} F_{qwe}(X - \rho E_i[Y^{qwe}_i]) dx dz
\]

Where \(F_{dj}\) denotes the cumulative distribution of non-pecuniary benefits in the population for occupation \(d\) and sector \(j\), with \(f_{dj}\) denoting the corresponding density function. For simplicity, \(f_{dse}(X - \rho E_i[Y^{dse}_i] * I(z \geq \frac{k_d}{\lambda^{d\beta}}))\) will from this point forward be denoted as \(f_{dse}\), while \(f_{dwe}(X - \rho E_i[Y^{dwe}_i])\) will be denoted as \(f_{dwe}\). The cumulative distribution functions will be denoted in a similar manner. I place no assumptions on cross-occupation or cross-sector interdependence of non-pecuniary benefits, and I place no restrictions on the functional form of non-pecuniary benefits other than assuming that \(\lim_{x \to -\infty} f_{dj}(x) = 0 \forall \ d, j\) and \(\lim_{x \to \infty} f_{dj}(x) = 0 \forall \ d, j\).

\(f(z)\) denotes the density function of asset characteristics \(q_i\), upon which I place no restrictive
assumptions, and $I(z \geq \frac{k_p}{\lambda^2\beta})$ is an indicator function that takes on the value of 1 if the individual possesses sufficient assets to meet the capital requirement of occupation $d$, and 0 otherwise.

Likewise, the probability that an individual chooses pair $dq, we$ is given by:

$$Pr(j = we & d = dq) = \int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{dqwe}(X - \rho \cdot E_i[Y_i^{dqwe}]) \prod_p F_{pse}(X - \rho \cdot E_i[Y_i^{pse}] \cdot I(z \geq \frac{k_p}{\lambda^2\beta})) \prod_{q \neq d} F_{qwe}(X - \rho \cdot E_i[Y_i^{qwe}]) dx dz$$

Conditional on working in an occupation $dq$, the probability that an individual is self-employed is given by:

$$Pr(j = se | d = dq) = \frac{Pr(j = se & d = dq)}{Pr(j = se & d = dq) + Pr(j = we & d = dq)}$$

For a given increase in asset values, denoted by an increase in $\beta$, the probability of self-employment conditional on working in occupation $dq$ changes as follows:

$$\frac{\delta Pr(j = se | d = dq)}{\delta \beta} = \frac{\int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{dqse}(X - \rho \cdot E_i[Y_i^{dqse}]) \prod_p F_{pse}(X - \rho \cdot E_i[Y_i^{pse}] \cdot I(z \geq \frac{k_p}{\lambda^2\beta})) \prod_{q \neq d} F_{qwe}(X - \rho \cdot E_i[Y_i^{qwe}]) dx dz}{(Pr(j = se & d = dq) + Pr(j = we & d = dq))^2}$$

Added up over all occupations $dq$, the total change in the probability that an individual is self-employed for a given occupational distribution is given by:

$$\sum_{dq=1}^{D} \int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{dqse}(X - \rho \cdot E_i[Y_i^{dqse}]) \prod_p F_{pse}(X - \rho \cdot E_i[Y_i^{pse}] \cdot I(z \geq \frac{k_p}{\lambda^2\beta})) \prod_{q \neq d} F_{qwe}(X - \rho \cdot E_i[Y_i^{qwe}]) dx dz$$

Where

$$\delta Pr(j = se | d = dq) = \int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{dqse}(X - \rho \cdot E_i[Y_i^{dqse}]) \prod_p F_{pse}(X - \rho \cdot E_i[Y_i^{pse}] \cdot I(z \geq \frac{k_p}{\lambda^2\beta})) \prod_{q \neq d} F_{qwe}(X - \rho \cdot E_i[Y_i^{qwe}]) dx dz$$

and

$$\delta Pr(j = we & d = dq) = \int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{dqwe}(X - \rho \cdot E_i[Y_i^{dqwe}]) \prod_p F_{pse}(X - \rho \cdot E_i[Y_i^{pse}] \cdot I(z \geq \frac{k_p}{\lambda^2\beta})) \prod_{q \neq d} F_{qwe}(X - \rho \cdot E_i[Y_i^{qwe}]) dx dz$$
Therefore

\[
\delta Pr(j=se \& d=d1) Pr(j = we \& d = d1) - Pr(j = se \& d = d1) \delta Pr(j=we \& d=d1) = \\
(f_{-\infty}^{\infty} f(z) f_{-\infty}^{\infty} \frac{\delta f_{\delta qse}}{\delta X} (-\rho \ast E_i[Y_1^{\delta qse}] \ast \frac{\Delta f(z \geq \frac{b_{dq}}{\Delta \beta}}{\Delta \beta}) \prod_{p \neq dq} F_{pse} \prod_{q} F_{qwe}dxdz \\
+ (f_{-\infty}^{\infty} f(z) f_{-\infty}^{\infty} f_{\delta qse} \sum_{p \neq d1} f_{pse} (-\rho \ast E_i[Y_1^{pse}] \ast \frac{\Delta f(z \geq \frac{b_{dp}}{\Delta \beta}}{\Delta \beta}) \prod_{l \neq p,dq} F_{ise} \prod_{q} F_{qwe}dxdz) * \\
(f_{-\infty}^{\infty} f(z) f_{-\infty}^{\infty} f_{\delta qwe} \prod_{p \neq d1} F_{pse} \prod_{q} F_{qwe}dxdz) \\
- (f_{-\infty}^{\infty} f(z) f_{-\infty}^{\infty} f_{\delta qwe} \sum_{p} f_{pse} (-\rho \ast E_i[Y_1^{pse}] \ast \frac{\Delta f(z \geq \frac{b_{dp}}{\Delta \beta}}{\Delta \beta}) \prod_{l \neq p} F_{ise} \prod_{q} F_{qwe}dxdz) * (f_{-\infty}^{\infty} f_{\delta qwe} \prod_{p \neq d1} F_{pse} \prod_{q} F_{qwe}dxdz) \\
\]

Using integration by parts on the first term in the above expression, with \(f'(x) = \frac{\delta f_{\delta qse}}{\delta X}\) and \(g(x) = (-\rho \ast E_i[Y_1^{\delta qse}] \ast \frac{\Delta f(z \geq \frac{b_{dq}}{\Delta \beta}}{\Delta \beta}) \prod_{p \neq dq} F_{pse} \prod_{q} F_{qwe}\) this term becomes:

\[
f_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f(x)g(x)dx = - (0 - f_{-\infty}^{\infty} f(z) f_{-\infty}^{\infty} f_{\delta qse} \prod_{p \neq dq} F_{pse} \prod_{q} F_{qwe}dxdz)
\]

The expression for \(\delta Pr(j=se \& d=dq) Pr(j = we \& d = dq) - Pr(j = se \& d = dq) \delta Pr(j=we \& d=dq)\) can then be simplified down to:

\[
\int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{\delta qse} (1-\rho \ast E_i[Y_1^{\delta qse}] \ast \frac{\Delta f(z \geq \frac{b_{dq}}{\Delta \beta}}{\Delta \beta}) \sum_{p \neq dq} f_{pse} (-\rho \ast E_i[Y_1^{pse}] \ast \frac{\Delta f(z \geq \frac{b_{dp}}{\Delta \beta}}{\Delta \beta}) - 1) \prod_{l \neq p,dq} F_{ise} \prod_{q} F_{qwe}dxdz \\
\int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{\delta qse} \prod_{p \neq dq} F_{pse} \prod_{q \neq dq} F_{qwe}dxdz \\
- \int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{\delta qse} \sum_{p} f_{pse} (-\rho \ast E_i[Y_1^{pse}] \ast \frac{\Delta f(z \geq \frac{b_{dp}}{\Delta \beta}}{\Delta \beta}) \prod_{q \neq dq} F_{qwe}dxdz & \\
\int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{\delta qse} \prod_{p \neq dq} F_{pse} \prod_{q \neq dq} F_{qwe}dxdz \\
- \int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{\delta qse} \sum_{p \neq dq} f_{qwe} \prod_{l \neq p} F_{ise} \prod_{q} F_{qwe}dxdz \\
\int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{\delta qse} \prod_{p \neq dq} F_{pse} \prod_{q \neq dq} F_{qwe}dxdz \\
\]

When this expression is added up across all possible occupations \(dq\), it becomes:

\[
\sum_{dq=1}^{D} \frac{\delta Pr(j=se \& d=dq)}{\delta \beta} Pr(j = we \& d = dq) - Pr(j = se \& d = dq) \frac{\delta Pr(j=we \& d=dq)}{\delta \beta} = \int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} f_{\delta qse} (\rho * 
\]

15
\[
E_i[Y_{dse}] = \theta(X_i, \eta_d)k_d^{\alpha_d} - r(M_i)k_d
\]
\[ \frac{\delta E_i[Y_{dq,se}]}{\delta \beta} = 0 \]

Therefore

\[ \sum_{d=1}^{D} \frac{\delta E_i[Y_{dq,se}]}{\delta \beta} = 0 \]

Since expected earnings conditional on being in a given occupation are not dependent upon \( \beta \), increases in asset values will not increase an individual’s expected earnings conditional on being self-employed for a given occupational distribution.

Therefore, based off of these results, when controlling for the occupational distribution, changes in \( \beta \), which I interpret as changes in local housing prices, are used as an exclusion restriction in a Heckman style model to predict the probability of self-employment without having an effect on earnings outcomes conditional on being self-employed in a given occupation.

5 Estimation

I expand upon the model from equation 1 by controlling for the difference between an individual’s expected earnings from self-employment and their expected earnings from wage employment.

2. \[ \Phi = \alpha_1 + \beta_x X_i + \beta_{imm} \cdot Imm_i + \beta_{\text{Difference}} \cdot \text{Difference}_i + \epsilon_i \]

Where \( \text{Difference}_i = E[Y_{dq,se}^i] - E[Y_{dq,we}^i] \)

An individual’s expected earnings from self-employment and from wage employment are calculated by estimating Heckman style models given by:

3. \[ \log(Y_{dq,we}^i) = \eta_{we} + \zeta_{we} + X_i + g_{we} S_{dq} + \rho_{we} \cdot \lambda_{we} + u_{we}^i \]

4. \[ \log(Y_{dq,se}^i) = \eta_{se} + \zeta_{se} + X_i + g_{se} S_{dq} + \rho_{se} \cdot \lambda_{se} + u_{se}^i \]
Where \( g_{we} \) and \( g_{se} \) represent the expected returns to the skills that are associated with an occupation in wage employment and self-employment respectively. \( \lambda_{we} \) and \( \lambda_{se} \) are the standard lambdas from the Heckman selection equation. I account for selection into self or wage employment by introducing unanticipated local housing price shocks (and it’s interaction with homeowner status) as a source exogenous variation to equation 1, with this equation being estimated jointly with equations 2 and 3 using Maximum Likelihood Estimation.

### 5.1 Local Housing Price Shocks

Housing collateral has been found to be an important source of small business financing (Adelino et al. 2015, Schmalz et al. 2017), and changes in housing value are anticipated to result in a higher propensity towards self-employment (Disney and Gathergood 2009). However, there are also concerns that changes in housing prices are related to local economic conditions, and therefore may possibly also be related to earnings potential in self or wage employment for reasons other than shocks in collateral availability. Previous work in the literature has used the residual component of housing price changes as a source of exogenous variation in the self-employment decision, with these residuals being cleaned of the effects of local macroeconomic characteristics such as the unemployment rate (Disney and Gathergood 2009). I use this approach in this paper.

I use county level data on housing prices and unemployment rates\(^{10}\). Furthermore, I use GDP per capita data at the Metropolitan Area level for urban areas, and the state level for rural areas\(^{11}\). I convert these variables to the PUMA level using the "MABLE/Geocorr14: Geographic Correspondence Engine" from the Missouri Census Data Center. Using data from 2001 to 2016, I construct the following fixed effects model to predict changes in local housing prices from lagged changes in unemployment rates and lagged GDP per capita:

\[
\%\Delta\text{HousingPrice}_{pu,t} = \sum_{k=0}^{k=7} \beta_{k1} \ast \Delta\text{UnemploymentRateChange}_{pu,t-k} + \sum_{k=0}^{k=7} \beta_{k2} \Delta\text{GDP Per Capita Change} + \alpha_{pu} + \gamma_t + \epsilon_{jt}
\]

\(^{10}\)The source for housing price data is the Federal Housing Finance Agency, with this data being available thanks to the work of Bogin et al. 2016. Unemployment rate data is taken from the Bureau of Labor Statistics.

\(^{11}\)GDP per capita data made available by the St. Louis Federal Reserve.
Where \( pu \) represents the PUMA and \( t \) represents the year. The dependent variable is the percentage change in local housing prices from the previous year, while the independent variables are lagged annual changes in the local unemployment rates and changes local GDP per capita going back to the previous 7 years\(^{12}\). \( \alpha_{pu} \) and \( \gamma_t \) capture PUMA and time specific fixed effects.

I include the residual component of the lagged annual change in housing prices over the course of the previous year in an individual’s PUMA as a source of exogenous variation in the Heckman equation. Furthermore, I include the interaction of these residual housing price shocks with a dummy variable indicating whether the individual has owned their current place of residence for at least the previous 2 years\(^{13}\). If the residual portion of housing price shocks effects the probability that one is self-employed through the channel of increasing the value of collateral, the results should show a positive effect of residual housing prices on self-employment probability for homeowners, but not for other individuals.

### 5.2 Skills

The U.S. Bureau of Labor Statistic’s O*NET database contains information on the characteristics that are expected of workers for each of the 840 occupations classified by the Standard Occupational Classification (SOC). I utilize the increasingly common approach among researchers\(^{14}\) of using the O*NET data in order to estimate an individual’s skill levels, under the assumption that workers are on average sorted into occupations that are a good match for their skills. The skill measures are constructed using the Skills, Knowledge, and Abilities content data contained in the O*NET database (hereafter collectively referred to as skills for simplicity). Each of these dimensions contain both importance and level scores\(^{15}\), which I then place on scales of 0-10.\(^{16}\) Since many of the

\(^{12}\)Since the 2008 financial crisis occurred during the sample period, I use 7 lags so that the predicted housing price changes for 2016 are calculated using post crash data.

\(^{13}\)I include only those individuals who have owned their residence for at least 2 years rather than using a contemporaneous homeownership variable only since the decision to purchase a home in an area in an area with a housing price shock is clearly not independent of one’s earnings. Therefore I exclude recent purchases in order to avoid this source of endogeneity.

\(^{14}\)Notable examples include Acemoglu and Autor (2011), Frey and Osborne (2017), and the pioneering work of Autor et al. (2003) using O*NET’s predecessor, the Dictionary of Occupational Titles.

\(^{15}\)Importance is a measure of the frequency with which that attribute is required on the job, while Level is a measure of the difficulty or intensity of the attribute that is required.

\(^{16}\)My choice of the Skills, Knowledge, and Abilities measures is largely influenced by Allen, Tsacoumis and McCloy (2011). The authors find that these measures are most closely related to an individual’s capacity to perform a job.
occupations that are included in the ACS data are aggregations of 2 or more of the detailed SOC occupations, I re-classify these aggregated occupations as the detailed occupation which is most common or representative of the group. In some cases the ACS data classifies an occupation as a miscellaneous member of a occupational category. For many of these cases the aggregation was judged to be too broad to assign the skills of any one occupation, therefore these groups are omitted from the estimation sample.

The ONET database contains measures of 109 different types of skills. However, it is clearly impractical to use them all in our analysis. Therefore, I create a series of orthogonal aggregations of these skill measures (including both the importance and level scores) using Principal Component Analysis, using the top 3 principal components in subsequent models. The first principal component captures primarily communication and language based skills, the second components captures primarily sensory and technical skills. The third principal component is heavily weighted towards levels of knowledge requirements. Together these three components account for approximately 62% of the variation in occupational skills.

6 Results

When estimating the selection equations for self-employment and wage employment (results in Tables 3 & 4), the two exclusion restrictions containing residual housing price shocks show that for non-homeowners, residual housing price shocks contain no explanatory power in terms of predicting an individual’s propensity towards self-employment. However, for homeowners of at least 2 years, an increase in residual housing prices is associated with a large and significant increase in self-employment probability. I also find that the housing price shocks have no effect on the probability that an individual is wage employed.

The earnings component of these models show that Component 1, the language based principal component, carries a significant earnings premium in both self and wage employment, with this

\footnote{A complete list of these re-classifications is available upon request.}

\footnote{An example of such a broad category is "Miscellaneous managers, including funeral service managers and postmasters and mail superintendents". With the exception of the aforementioned occupation, representing 3.23% of the total data, each of these occupations comprise less than 0.1% of the total observations. A list of these omitted occupational groups is available upon request.}
Table 2: Estimated Effects of Principal Components on Earnings in the Self-Employment and Wage Employment Sectors

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.076***</td>
<td>0.011***</td>
<td>-0.019***</td>
</tr>
<tr>
<td>-0.075**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lambda</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.091***</td>
<td>0.06***</td>
<td>-0.027***</td>
</tr>
<tr>
<td>-0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lambda</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.02***</td>
<td>-0.04***</td>
<td>0.037***</td>
</tr>
<tr>
<td></td>
<td>Homeowner</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td></td>
</tr>
<tr>
<td>-0.023***</td>
<td>0.049</td>
<td>0.687***</td>
</tr>
<tr>
<td>Homeowner*Residual</td>
<td>0.353</td>
<td>0.289</td>
</tr>
</tbody>
</table>

Table 3: Housing Price Shocks and Self-Employment Propensity

<table>
<thead>
<tr>
<th>Component 1</th>
<th>Component 2</th>
<th>Component 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.056***</td>
<td>0.022***</td>
<td>-0.047***</td>
</tr>
<tr>
<td>Homeowner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.023***</td>
<td>0.353</td>
<td>-0.217</td>
</tr>
<tr>
<td>Homeowner*Residual</td>
<td>0.289</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Housing Price Shocks and Wage Employment Propensity

6.1 Effect on the Probability of Self-Employment

After estimating an individual’s predicted earnings potential in both self-employment and wage employment, I then estimate the difference between these estimated earnings potentials according to

premium being slightly larger for the self-employed. I also find a small penalty associated with Component 3 in both sectors. The most notable difference is in the returns to the second, manual, principal component. Manual skills are estimated as carrying a significant premium in the wage sector, compared to a small premium in the self-employment sector. The results from these equations also show that individuals with a lower estimated probability of being self-employed tend to earn less conditional on being self-employed.

6.1 Effect on the Probability of Self-Employment

After estimating an individual’s predicted earnings potential in both self-employment and wage employment, I then estimate the difference between these estimated earnings potentials according to

premium being slightly larger for the self-employed. I also find a small penalty associated with Component 3 in both sectors. The most notable difference is in the returns to the second, manual, principal component. Manual skills are estimated as carrying a significant premium in the wage sector, compared to a small premium in the self-employment sector. The results from these equations also show that individuals with a lower estimated probability of being self-employed tend to earn less conditional on being self-employed.

6.1 Effect on the Probability of Self-Employment

After estimating an individual’s predicted earnings potential in both self-employment and wage employment, I then estimate the difference between these estimated earnings potentials according to

premium being slightly larger for the self-employed. I also find a small penalty associated with Component 3 in both sectors. The most notable difference is in the returns to the second, manual, principal component. Manual skills are estimated as carrying a significant premium in the wage sector, compared to a small premium in the self-employment sector. The results from these equations also show that individuals with a lower estimated probability of being self-employed tend to earn less conditional on being self-employed.

6.1 Effect on the Probability of Self-Employment

After estimating an individual’s predicted earnings potential in both self-employment and wage employment, I then estimate the difference between these estimated earnings potentials according to

premium being slightly larger for the self-employed. I also find a small penalty associated with Component 3 in both sectors. The most notable difference is in the returns to the second, manual, principal component. Manual skills are estimated as carrying a significant premium in the wage sector, compared to a small premium in the self-employment sector. The results from these equations also show that individuals with a lower estimated probability of being self-employed tend to earn less conditional on being self-employed.
Table 5: Effect of Differences in Earnings Potential on Self-Employment Propensity

<table>
<thead>
<tr>
<th>Equation Type</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Coefficient for Immigrants and Natives Difference</td>
<td>0.04***</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Varying Coefficients for Immigrants and Natives Difference</td>
<td>0.037***</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Imm*Difference</td>
<td>0.024***</td>
<td>(0.004)</td>
</tr>
</tbody>
</table>

The literature on immigrant self-employment has primarily been divided into two opposing camps, with the "push" theory of immigrant self-employment claiming that immigrants are more likely to be self-employed in large part because of poor opportunities in the wage employment sector. This stands in contrast to the "pull" theory of self-employment which contends that immigrants tend to possess stronger abilities in self-employment relative to natives, and that they respond by choosing
The results of this paper suggest, in contrast to these two theories, that immigrant’s higher propensity towards self-employment does not stem from differences in earnings potential in self-employment relative to natives, but rather that conditional on occupational skills, immigrants are more likely to choose the sector in those skills are more highly valued.

8 References


Heckman, James J. ”Sample selection bias as a specification error (with an application to the estimation of labor supply functions).” (1977).

Heim, Bradley T., and Ithai Z. Lurie. ”The effect of self-employed health insurance subsidies on


