NOVAFRICA
Working Paper Series

Occupation-Skill Mismatch and Selection of Immigrants: Evidence from the Portuguese Labor Market

Tijan L. Bah
Nova School of Business and Economics (Universidade Nova de Lisboa), NOVAFRICA, University of Paris 1 Panthéon-Sorbonne, and IRD, DIAL

ISSN 2183-0843
Working Paper No 1804
December 2018

Occupation-Skill Mismatch and Selection of Immigrants: Evidence from the Portuguese Labor Market*

Tijan L Bah†

December 2018

Abstract

This paper aims at investigating how the occupational placement of immigrants relative to their qualifications affect their self-selection. Using an administrative matched employer-employee data set for Portugal for the years 2002-2009, we first estimate the probability that an average worker from a particular country is overeducated, matched, or undereducated relative to the skill needs of the occupation he takes upon immigration. Second, using these estimated probabilities, we analyze how overeducation and appropriate skill-occupation matches affect selection of immigrants from 40 origin countries. The results suggest that overeducation leads to negative self-selection of immigrants into the Portuguese labor market. Furthermore, the evidence suggests that appropriate occupation-skill matches affect migration selection positively. These results imply that receiving countries' selective policies aimed at attracting high skilled immigrants should also focus on reducing occupation-skill mismatch probably through degree recognition and standardization in collaboration with sending countries.

Keywords: Selection, Occupation-Skill Mismatch, Portugal, Immigrants, Waste, Immigration.

JEL Codes: F22, J24.

---

*I am grateful to Catia Batista, Pedro Freitas, Marta C. Lopes and Christoph Deuster for their helpful comments and suggestions. I also thank participants of the 2016 Nova SBE PhD Research Group Meeting, 14th IMISCOE Annual Conference, 10th Annual Meeting of the Portuguese Economic Journal and IX Italian Doctoral Workshop in Empirical Economics. I am grateful to the Ministry of Employment for providing the data used and to Nova Fórum and Fundação para a Ciência e Tecnologia for financial support. I acknowledge the European Doctorate Degree in Economics - Erasmus Mundus (EDE-EM) Program for funding and support.

†Nova School of Business and Economics - Universidade Nova de Lisboa, NOVAFRICA, University of Paris 1 Panthéon-Sorbonne, and IRD, DIAL. Email: tijan.bah@novasbe.pt

1 Introduction

Over the last 15 years, the number of international migrants has grown rapidly. The stock of international migrants reached 244 million in 2015 compared to 222 million in 2010, 191 million in 2005 and 173 million in 2000 (United Nations, 2016). The effect of this growing stock of migrants on both receiving and sending countries has been the object of growing interest by both the academic community and policymakers.

In order to better understand the consequences of international migration, several important questions arise: What are the characteristics of migrants? What is the quality of human capital migrants possess? Is this capital adequately valued in the international labor market? Though there is consensus on some of the answers to these questions, others continue to be controversial.

A big debate remains on the characteristics of international migrants. These characteristics include both observables such as age group, gender structure, education levels, and non-observable characteristics such as ability, risk and discount preferences. There are numerous studies examining the skills or educational levels of those who migrate. Most of the debate has been centered on the characteristics or self-selection of immigrants especially from Mexico to the United States. To date, there is no consensus on the selection of immigrants. On the one hand, studies such as Chiquiar and Hanson (2005), Orrenius and Zavodny (2005), and McKenzie and Rapoport (2010) finds positive selection of immigrants from Mexico to the United States. On the other hand, Borjas (1987), Moraga (2011), and Ambrosini and Peri (2012) concludes negative selection of immigrants.

It is well documented that there is less than perfect international transferability of immigrants' human capital into foreign labor markets, especially for the high skilled (Chiswick and Miller, 2009; Kiker et al., 1997; Mattoo et al., 2008). The skills that immigrants acquire before migration do not seem to be fully utilized after international migration. In particular when immigrants are compared to natives, the evidence reveals that within the same occupation, immigrants are mostly overeducated (brain waste or occupation-skill mismatch). This occupation-skill mismatch has negative impacts on the returns to investment in education (Alba-Ramírez and Segundo, 1995; Carneiro et al., 2012; Duncan and Hoffman, 1981; Kiker and Santos, 1991; Verdugo and Verdugo, 1989).

The combination of these pieces of evidence raises two important questions. First: Does imperfect international transferability of human capital affects the decision to migrate? According to Sjaastad (1962), migration is an investment in which individuals engage in migration when the expected return exceeds the expected cost of migrating. The expected returns crucially depend on wages earned in the destination country and in turn wages are a function of the skills possessed and their market value. This very simple model is enough to justify the above research question. An important second research question is: Does the imperfect international transferability of human capital affects the migration decision of the high skilled and the low skilled differently? The answer to this research question has important policy implications in that it can affect "brain drain" flows, which have been shown to be very relevant for the economic performance of both origin and destination countries of migration[1].

This paper aims at examining precisely how the self-selection of immigrants may be affected by existing occupation-skill mismatches. This is a novel research question relative to the existing literature. While the effects of occupation-skill mismatches on wages have been well documented in the literature[2], little is known about whether and how occupation-skill mismatch affects who migrates. Most of earlier research explains the selection of immigrants with origin country earnings inequality, costs of migration, differences between origin and destination skill earnings premium (Batista and McKenzie, 2018; Borjas, 1987; Chiquiar and Hanson, 2005; Grogger and Hanson, 2011), and recently migration networks and diaspora (Beine et al., 2011; McKenzie and Rapoport, 2010). We argue and provide supportive evidence that the selection of immigrants can also be explained by occupation-skill mismatch.

The main aim of this paper is therefore to determine whether imperfect transferability of human capital or occupation-skill mismatch affects selection of immigrants. To achieve this objective, we first augment the model of migration proposed in Beine et al. (2011) by including occupation-skill mismatch as another determinant of migration.

Second, using a Portuguese matched employee-employer panel data for the years 2002-2009, we compute the probability that an average worker from a particular country is undereducated, or correctly matched, or overeducated. These probabilities reveal that immigrants - especially those who are high skilled, i.e. immigrants with post secondary education - are more likely to be overeducated and less likely to be correctly matched to an occupation than natives.

Finally, we analyze empirically how occupation-skill mismatch affects the selection of immigrants. We find that high skilled migrants are less likely to migrate than the low skilled when the difference between the probability of overeducation for the high skilled and the low skilled increases. Furthermore, we show that when the probability of correct skill match increases, the high skilled are more likely to migrate compared to the low skilled.

The remainder of the paper is organized as follows. In the next section, we survey related literature. Section 3 discusses the theoretical model, the data sources and the empirical estimation. In section 5, we discuss the main results of the paper. Section 6 concludes and gives some policy implications.

[1] This discussion is well informed by the traditional references such as Bhagwati and Hamada (1974) or Grubel and Scott (1966) on the losses implied by the brain drain for the migrant home countries. This negative view has however been balanced by more recent literature emphasizing the potential benefits of skilled migration to the country of origin, namely in terms of educational gains such as shown by Beine et al. (2001, 2008) or Batista et al. (2012), institutional improvements such as Batista and Vicente (2011) or Batista et al. (2016), entrepreneurial gains and business investment such as Yang (2008) or Batista et al. (2017).

[2] See Hartog (2000) for a meta analysis of occupation-skill mismatch and earnings.

2 Related Literature

Since the famous work of Roy (1951) model of selection, understanding selection of immigrants has been in the forefront of research. The first work that formalizes selection in the context of migration is the influential work of Borjas (1987) who applies the Roy (1951) model to analyze the selection of immigrants from Mexico to the United states. According to Borjas (1987) theoretical model's prediction, selection of immigrants depends crucially on the earnings inequality of the origin and destination countries. In particular, immigrants from a country with high earning inequality to a country with low earning inequality will be negatively selected, that is migrants are drawn from the lower half of the skill distribution. Borjas (ibid.) model, though influential, has been criticized because of its assumption that migration cost is constant across individuals.

Chiquiar and Hanson (2005) argue that selection of immigrants can be positive or negative depending on the cost of migration and its heterogeneity across skill groups. The costs of migration for high skilled is likely to be lower than the low skilled. Their work compares Mexicans who migrated to the United States and those who stayed using census data of both Mexico and the United States. In a similar spirit, McKenzie and Rapoport (2010) show that migration networks play a crucial role in determining the selection of immigrants. In particular, they showed that migrants from communities with high migration are negatively selected while those from communities without established migration networks are positively selected. They used data from the Mexican Population Project (MPP). This result confirms the important role of migration costs as networks are expected to reduce the cost of migration.

Recently, Moraga (2011) and Ambrosini and Peri (2012) showed results of negative selection as predicted by Borjas (1987). The debate on the selection of immigrants especially from Mexico to the United States remains controversial as different authors conclude differently depending the data set used.

Elsewhere, the selection of immigrants to Europe and other countries has been less studied compared to the USA. However, recently it has attracted some attention probably because of the increase migration flows over the last couple of decades. Beine et al. (2011) studied how diaspora (stock of migrants) affects the selection of immigrants into 30 OECD countries in 1990 and 2002. Their analysis reveals that diaspora lowers migration costs and thus increases migration flows especially the low skilled and hence reduces the quality of immigrants in terms of educational attainment. Similarly, Grogger and Hanson (2011) studied the selection and sorting of immigrants in the OECD countries. They concluded that immigrants are positively selected if the absolute skill-related difference in earnings between the destination country and the source is large.

Quinn and Rubb (2005) first studied how education-occupation matches influence the migration decisions of Mexicans to the United States. Backed by a simple theoretical model, they showed that indeed the decision to migrate is influenced not only by the actual years of education but also both by the years of overeducation and undereducation. The results suggest that Mexican workers who are overeducated are more likely to migrate while undereducated workers are less likely to migrate. They argued that the desire to migrate by overeducated workers is driven by the potential for better matches and subsequent increase in earnings.

This work is broadly related to the brain drain literature. On the one hand, the brain drain literature such as argues about the negative welfare effects of high skilled emigration on those left behind (Bhagwati and Hamada, 1974), while Haque and Kim (1995) showed that brain drain has negative effects on economic growth of sending countries. On the other hand, the beneficial brain drain literature argues that migration prospects under uncertainty have a positive impact on the human capital development of the country of origin. The work of Beine et al. (2001, 2008) offered important theoretical and empirical contributions of the macroeconomic implications of brain drain. Beine et al. (2001) theorized that the prospects of international migration motivates investment in education. Beine et al. (2008) offers empirical evidence that indeed high skilled emigration has positive benefits on human capital development of some source countries while negative effect for others.

The beneficial brain drain literature has been criticized for being overly optimistic. The impact of brain drain on growth and welfare are greatly exaggerated (Schiff, 2005). His theoretical analysis asserts that allowing for brain waste possibility and factoring out the additional cost of investment in education instead of investing in public good provisions will result in smaller or zero gain in economic growth and welfare. Along similar reasoning, Pires (2015) criticized the model for not considering the role of brain waste or occupation-skill mismatch. He theoretically argued that brain waste reduces the incentive to invest in education, lowers the chances of positive self-selection, and thus reduces the possibility of beneficial brain gain.

The literature on selection of immigrants though vast have neglected an important phenomenon of migration labour market; occupation-skill mismatch also known as brain waste. Though the implications of skill mismatch on wages has been documented, with the exception of Pires (ibid.), who models how the risk of brain waste affects brain drain; to the best of my knowledge, the impacts of mismatch on selection of immigrants haven't been studied. This paper aims at filling this void in the literature by introducing occupation-skill mismatch as a factor that affects the selection of immigrants.

3 Methodology

3.1 Theoretical Framework

Following the framework developed by Beine et al. (2011) and Grogger and Hanson (2011), consider workers decision to migrate governs by a simple linear utility maximization problem. Specifically, assume that utility of an individual with skill level h born in country i and staying in country i is given by:

$$U_{ii}^h = w_i^h + A_i + \epsilon_i, \tag{1}$$

where w_i is the wage in country i, A_i is country i*'s* characteristics and ϵ_i is an iid extreme value distributed error term. If an individual migrated to another country j, the utility is given by:

$$U^h_{ij} = w^h_{ij}(M^h_{ij}) + A_{ij} + M^h_{ij} - C^h_{ij} + \epsilon_j, \tag{2}$$

where M^h_{ij} is the quality of the skill-job match of an immigrant from country i in country j. We assume that the skill gathered by the worker before migrating might not be fully valued in the destination country. As shown in the literature, we also assume that wages are affected by the quality of the match. Moreover, we assume that individuals utility is directly affected by the quality of the skill-job match. In addition, C^h_{ij} denotes the cost of migrating from country i to country j.

Denote N_i as the number of people within migration age in country i and let N_{ij} denote the number of people who actually migrated to country j. Under the assumptions that errors follow iid extreme value distribution, as shown by McFadden (1984), the log odds of skilled level h group from country i migrating to country j can be written as,

$$\left[\frac{N^h_{ij}}{N^h_{ii}}\right] = (w^h_{ij}(M^h_{ij}) - w^h_i) + (A_{ij} - A_i) + M^h_{ij} - C^h_{ij} \tag{3}$$

Equation 3 implies that odds ratio of group h migrating depends on the difference between destination wage (which is a function of the skill-job match) and the origin wage. The odds ratio also depends directly on the skill-job match, the cost of migrating, and the difference between the destination and the origin country characteristics.

Consider individuals endowed with two types of skills, either high skill or low skilled. Let $h = 0$ and $h = 1$ for low skilled and high skilled respectively. Thus we can write the logs odds of migrating from country i to country j as:

$$ln\left[\frac{N^0_{ij}}{N^0_{ii}}\right] = (w^0_{ij}(M^0_{ij}) - w^0_i) + (A_{ij} - A_i) + M^0_{ij} - C^0_{ij} \tag{4}$$

$$ln\left[\frac{N^1_{ij}}{N^1_{ii}}\right] = (w^1_{ij}(M^1_{ij}) - w^1_i) + (A_{ij} - A_i) + M^1_{ij} - C^1_{ij} \tag{5}$$

Following Grogger and Hanson (2011), using equations 4 and 5 we can analyze the selection of immigrants to country j. Taking the difference between equations 4 and 5 and rearranging, we obtain:

$$ln\left[\frac{N^1_{ij}}{N^0_{ij}}\right] = \left[w^1_{ij}(M^1_{ij}) - w^0_{ij}(M^0_{ij})\right] + \left(M^1_{ij} - M^0_{ij}\right) - \left(C^1_{ij} - C^0_{ij}\right) + ln\left(N^1_{ii}/N^0_{ii}\right) - \left(w^1_i - w^0_i\right) \tag{6}$$

Equation 6 indicates that selection of migrants is determined by five terms. The first term gives the difference between the earnings of the two groups. The second term is the difference between the quality of occupation-skill match. While the third term gives the difference between the cost of migrating for the high skilled and low skilled individuals. The final two terms of equation 6 are the log ratio of population of high skilled to low skilled of country i and the differences of earnings of high skilled and low skilled in the origin country. We expect that there will be positive selection, that is the high skilled are more likely to migrate, if wage gap in the destination country is positive. Moreover, there will be positive selection if skills of the high skilled group are more valued in the destination country relative to the low skilled.

3.2 Data and Descriptive Statistics

The primary data is from the *Quadros de Pessoal (QP)*. The QP is a rich data set of firms and employees collected yearly by the Portuguese Ministry of Labour and Social Solidarity. All firms that have at least 10 employees are mandatory required by law to provide data about its employees and other firm characteristics. The data of the employees includes worker unique social security identifier, gender, age, tenure, education, earnings, hours worked, etc. The firm characteristics includes location,economic activity and number of employees.

The data cover years from 1986 to date with the exception of 1990 and 2001. We restrict our analyses for periods 2002 to 2009 as the nationality of workers were not documented before 2000. The data set has no information on year of migration of the worker. However, following Cabral and Duarte (2013), we proxy the year of migration with the year the worker first appeared in the database.

The education years of workers are coded as: 0 years (less than 1st cycle) , 4 years (1st cycle) , 6 years (2nd cycle), 9 years (3rd cycle), 12 years (secondary), 13 years (post secondary but less than *Bacharelato*), 14 years (*Bacharelato*), 15 years (*Licenciatura*), 17 years (master), and 20 years (doctorate). In line with Beine et al. (2011) and Grogger and Hanson (2011), we define the high skilled as workers with those more than secondary level of education. Experience is defined as age less tenure less years of education and 6 years . In addition, the occupational category used is the 3 digit *Classificação National de Profissões* (CNP 94) which have a total of 276 occupations.

We made some restrictions in the data set. First, for workers who appear more than once in the data set in the same year, we keep the job with the highest wage. Second, we restricted the sample to workers with ages between 15 years and 80 years. Moreover, since it is illegal by law to pay workers below 80 per cent of the minimum wage (Cabral and Duarte, 2013), only workers with reported of earnings more than the 80 per cent of the 2002 minimum wage are included in the analyses. Finally, the analysis is restricted to the top 40 sending countries of migrants.

Table 1 below presents descriptive statistics of selected variables by skill group. The proportion of high skilled immigrants accounts for 30 per cent of the sample during the years 2002 to 2009. On average, high skilled immigrants have 14.10 years of education while low unskilled have 6.19 years of education. Moreover, high skilled have a higher proportion of females, and lower wages compared to the low skilled group. However, the low skilled immigrants have on average higher years of experience, hours of work, and years in Portugal compared to high skilled immigrants.

Additional data on GDP per capita and population is obtained from the World Bank development indicators and United Nations Population Division, respectively.

Table 1: Sample Means of Selected Variables

Variable	Full Sample	High skilled	Low skilled	Difference
High skilled	0.36	-	-	-
Age	35.49	34.75	35.89	-1.13***
Education	9.64	14.10	7.19	-6.19***
Tenure	3.49	2.01	2.10	-0.86***
Experience	15.54	12.23	18.69	-6.46***
ln(Wages)	6.51	6.61	6.49	0.14***
Female	0.34	0.35	0.33	-0.02***
Years since migration	3.54	3.01	3.83	-0.82***
Observations	793,169	280,973	512,196	-

Source: *Quadros de Pessoal* database 2002-2009.
* p<0.1; ** p<0.05; *** p<0.01.

3.3 Occupation-Skill Mismatch

3.3.1 Mismatch and its Determinants

In this paper, two methods of computing the occupation-skill mismatch are employed. The first method following Verdugo and Verdugo (1989) requires computing the mean and standard deviation of the education for each occupation. Under this method, overeducated (undereducated) workers are those with education level more (less) than one standard deviation of the mean for their occupation. The second method first proposed by Kiker et al. (1997) computes mismatch by using the modal years of education for each occupation. Workers are classified as overeducated (undereducated) if their education attainment is more (less) than modal education attainment of their occupation.

Following these methods, we compute occupation-skill mismatch using the three digit *Classificação National de Profissões (CNP 94)*. The CNP is a standard detailed classification of occupations which contains four digits. The third digit classification employed in the analyses has 276 occupations. As argued by Tsai (2010), in order to minimize measurement error in the computation of mismatch, I assume sector heterogeneity in valuating education. Therefore,for each year, the required education attainment is computed by occupation and by sector. Moreover, the total sample of both natives and migrants are used to compute the mismatch.

Table 2 presents the incidence of mismatch disaggregated by skills level. The estimates shows that the incidence of overeducation for the full sample ranges from 52 % for the modal procedure to 33 % from the range method. It is also observed that high skill workers are less likely to be correctly matched than the low skilled. In order to compute the probability of these mismatches by country and year, multinomial logit model is employed. This model can be expressed as:

$$Y_{ij}|X_i = \frac{e^{\beta_j X_i}}{\sum_{k=1}^{3} e^{\beta_k X_i}}; i = 1, 2..., n; j = 1, 2, 3. \quad (7)$$

Table 2: Incidence of Mismatch by skill group in 2002-2009

Variable	Overeducated	Matched	Undereducated	Observations
Mode				
Low Skilled	38.24	38.96	22.88	512,196
High Skilled	80.72	14.60	4.68	280,973
Full Sample	53.29	30.33	16.38	793,169
Range				
Low Skilled	18.27	65.50	16.23	512,196
High Skilled	61.05	38.24	0.71	280,973
Full Sample	33.42	55.84	10.24	793,169

Source: *Quadros de Pessoal* database 2002-2009 and own calculation.

Where Y takes values 1 if the worker is overeducated, 2 if correctly matched, and 3 if undereducated. X is a vector of worker characteristics which includes job tenure, experience, gender, nationality, years since migration, and skill group. After estimating the multinomial logit, the probabilities that an average worker belonging to the three categories can easily be predicted for each year and country and for each skill group.

3.4 The Effects of Mismatch on Selection of Immigrants

From the theoretical section, we observed that selection of immigrants is determined by wages, occupation skill mismatch, and costs of migration. We can use equation 6 to analyze the effects of occupation-skill mismatch on the selection of immigrants by using the following econometric specification:

$$lnS_{it} = \gamma lnS_{it-1} + \alpha_0 OEdiff_{it} + \alpha_1 Wagediff_{it} + \beta' X_{it} + \eta_i + \mu_t + \epsilon_{it} \tag{8}$$

Where S_{it} is the ratio of high skilled to low skilled flows of migrants from country i in time t, $OEdiff$ is the difference between the probabilities of overeducation for high skilled and low skilled immigrants computed from the multinomial regression estimates. $Wagediff$ is hourly wage difference between high skilled and low skilled immigrants in Portugal. It is worth mentioning that, unlike Chiquiar and Hanson (2005), due to data limitations, wages at the origin is excluded from the analysis. X is a vector of country characteristics that affects migration flows such as GDP per capita and population. While η_i and μ_t are country fixed effects and time fixed effects, respectively.

To estimate the above specification, we employ the system generalized method of moments (SGMM) proposed by Blundell and Bond (1998). Unlike the Arellano and Bond (1991) first-differenced generalized method of moments (DGMM) estimator, the SGMM estimation technique uses both the level equation of equation 8 and its first difference Blundell and Bond (1998). According to Blundell and Bond (ibid.), compared to the DGMM estimator, the SGMM estimator improves efficiency especially if the dependent variable is highly persistent and when the variance of the unobserved individual heterogeneity is high. Both the SGMM and DGMM estimator rely on the assumption that the first difference errors are autoregressive of order one.

Throughout the estimation, we assume that the variable lagged diaspora is predetermined while the rest of the variables in equation 9 are endogeneous; thus requiring instruments for these variables. These instruments are obtained by using the lags of the variables in the model. Fortunately, the validity of these instruments can be tested using the Sargan overidentifying restriction test.

4 Results

4.1 Occupation-Skill Mismatch and its Determinants

To compute the probabilities of occupation-skill mismatch, the results from the multinomial logit model are analyzed. Table 3 shows the result from the multinomial logit regression using the modal definition of mismatch. The coefficients give how each of the variables affect the probability of overeducated or undereducated compared to being correctly matched. Column 1 of table 3 shows that log odds of being overeducated compared to being correctly matched is a decreasing function of years of experience and tenure. The coefficient on experience shows that older workers in the labour market performs relatively better in terms of the probability of being correctly matched. We also observe that high skilled and female workers are more likely to be overeducated. Surprisingly, the coefficient on years since migration is positive and statistically significant, implying that workers who spent more years in the country are more likely to be overeducated. One possible explanation for this is that immigrants augment their education level after migration thus making them more skilled and thus more likely to be overeducated.

On the other hand, as shown in column 2 of table 3, the odds of being undereducated increases with years of experience and tenure. The dummies on female and high skilled indicate that female and high skilled workers are less likely to be undereducated. Moreover, the years since migration reduce the probability of being undereducated.

Though the individual coefficient estimates from the multinomial logit model are somewhat less informative, the model allows us to predict the probability that an average worker from a particular country is overeducated, undereducated, or correctly matched. Moreover, we can further compute these probabilities by skill group for each country. These estimates are very accurate especially for the average worker. In order to predict these probabilities, averages of the variables in table 3 is employed for each county. Figure 1 shows the predicted probabilities by country that an average worker is overeducated using the modal definition in the year 2002.

The results from figure 1 shows that the probability of overeducation is heterogeneous across countries with Philippines recording the highest probability of 67% and China the lowest probability of 17%. One possible explanation for better matches for Chinese immigrants is due to them being highly entrepreneurial as shown in Oliveira (2003). As expected, immigrants from Portuguese speaking countries perform relatively better with Brazil recording the highest probability of 46%. Countries of the Schengen area also

Table 3: Multinomial Logit Model Estimates of Occupation-skill Mismatch

Variable	**Ln(OE/M)**	**Ln(UE/M)**
Experience	-0.0304***	0.0731***
	(0.0004)	(0.0004)
Experience square	0.000003	0.00001***
	(0.000003)	(0.000004)
Tenure	-0.0590***	0.0953***
	(0.0013)	(0.0012)
Tenure square	0.00001***	-0.00002***
	(0.000003)	(0.000004)
Female	0.0279***	-0.0290***
	(0.0075)	(0.010)
High skilled	2.5227***	-2.3963***
	(0.0087)	(0.0242)
Years since migration	0.0501***	-0.0816***
	(0.0022)	(0.0025)
Years since migration squared	-0.0035	0.0023***
	(0.0001)	(0.0001)
Observations	703,864	
Pseudo R^2	0.1816	
Log likelihood	-512704.29	
χ^2 (154)	227472.40	

Source: *Quadros de Pessoal* database 2002-2009
Notes: Other controls not reported includes country and time dummies, firm size,location, and sector of activity. Standard errors in parenthesis. * $p < 0.1$; ** $p< 0.05$; *** $p < 0.01$.

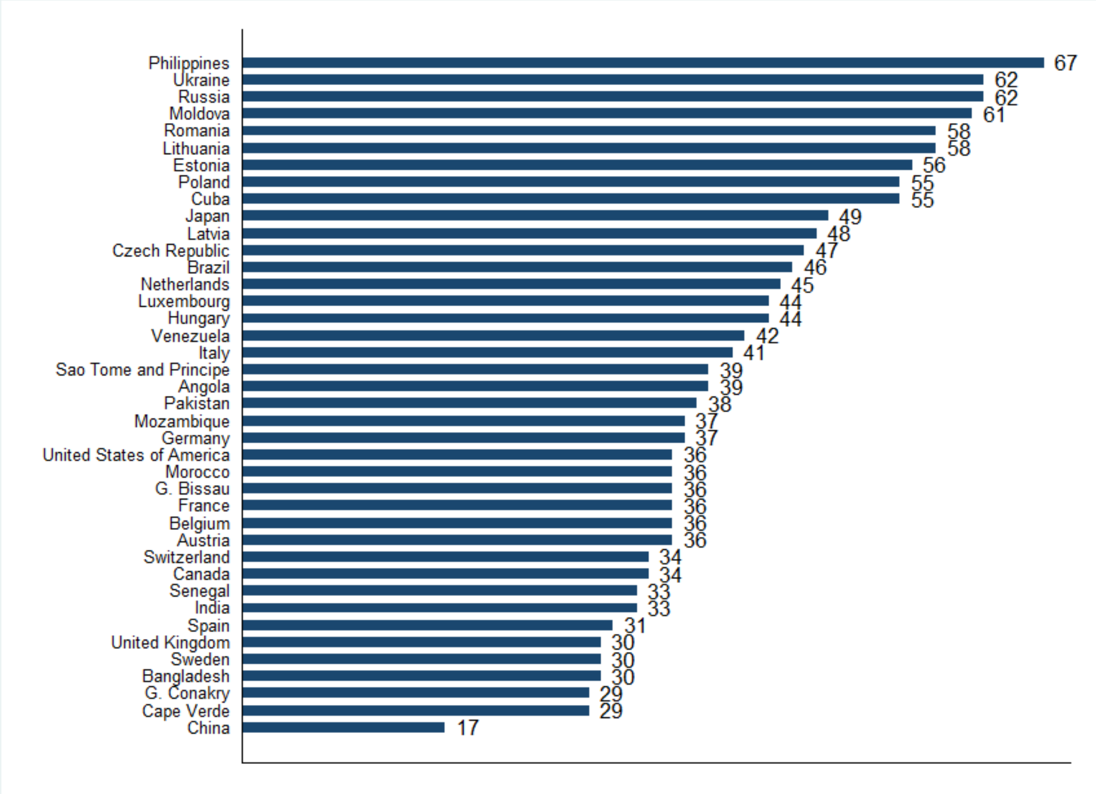
perform relatively better than the rest of Eastern Europe. Ukraine though the third largest contributor to the stock of immigrants after Brazil and Cape Verde records the second highest probability of overeducation. The results from figure 1 are consistent with Mattoo et al. (2008), who also demonstrated heterogeneous mismatch among immigrants in the United States.

4.2 Mismatch and Selection of Immigrants

Table 4 shows how occupation skill mismatch affects selection of immigrants. As mentioned earlier, both definitions of mismatch (range and mode) are employed. However, irrespective of the definition, we observe consistent negative effect of mismatch on selection.

Specifically, column 1 of table 4 shows how mismatch affects selection using the range definition. We observe that the coefficient of overeducation differences as expected is negative and statistically significant at 5%. These results suggest that when the

Figure 1: Estimated probability (in percentages) that a worker is overeducated in 2002



probability of overeducation of the high skilled is greater than the low skilled, the high skilled are more likely to migrate compared to the low skilled (negative selection). It can also be observed that column 2 of table 4 shows similar findings though with a coefficient of higher magnitude. These results are consistent with our theoretical predictions that indeed selection of immigrants is also explained by skill mismatch.

The control variables included in the estimation show the expected signs with the exception of coefficient on differences between the wages of high skilled and low skilled. The potential reason for the negative sign of the coefficient is due to omitting the wages at the origin country. As shown in Grogger and Hanson (2011), estimating the selection equation using origin country wages is crucial. However, in this paper, we do not have data on wages in the origin countries. As observed elsewhere in Beine et al. (2011), diaspora leads to negative selection of immigrants. The reason for this negative effect stems from the established fact that diaspora provides networks assistance that tend to reduce the cost of migration especially for the unskilled. Moreover, from the coefficient of GDP per capita, the results suggest that immigrants from richer countries are positively selected.

As per the GMM assumptions of no serial correlation, the results suggest that indeed this assumption is satisfied. Moreover, the Sargan test shows that the overidentifying

Table 4: Impact of Overeducation on Selection

Variable	**Ln(skill ratio)** *Range Definition*	**Ln(skill ratio)** *Mode Definition*
Overeducation differences	-0.0135**	-0.0264**
	(0.0067)	(0.0117)
Wage differences	-0.0086	-0.0000
	(0.0166)	(0.0156)
Lagged diaspora (stock)	-0.0090	-0.0152*
	(0.0083)	(0.0085)
GDP Per Capita	0.0298***	0.0325***
	(0.0054)	(0.0043)
Population	-0.0006*	0.0000
	(0.0003)	(0.0003)
Lagged skill ratio	0.1533**	0.1161**
	(0.0580)	(0.0606)
Sargan P-Value	0.1420	0.2671
AR(1)	0.0002	0.0004
AR(2)	0.6817	0.9126
No of Instruments	119	117
Countries	40	40
Observations	262	262

* p<0.1; ** p<0.05; *** p<0.01

assumption is also satisfied with p-values of 0.1286 and 0.1063 for the range and mode definitions respectively.

Similarly, table 5 shows how correct matches affects selection. The results from 5 are also consistent with our theoretical predictions. The coefficient of occupation-skill match differences between high skilled and low skilled is as expected positive and statistically significant. This implies that when probability of correct occupation-skill match of the high skilled is higher than the low skilled, the high skilled are more likely to migrate; hence positive selection. Similar to results in table 4, with the exception of wages, the coefficients of the control variables have the expected signs.

The results from the post estimations show that the GMM assumptions are satisfied. As expected, the results from the autocorrelation test show that we cannot reject the hypothesis of zero autocorrelation in first-differenced errors of order 2. Moreover, the results from Sargan overidentifying restrictions test shows that moment conditions are satisfied with p-values of 0.1168 for the range definition and 0.1615 for the mode definition.

Table 5: Impact of Correct Match on Selection

Variable	**Ln(skill ratio)** *Range Definition*	**Ln(skill ratio)** *Mode Definition*
Occupation-skill match differences	0.0167**	0.0276**
	(0.0074)	(0.0187)
Wage differences	-0.0111	0.0001
	(0.0131)	(0.0156)
Lagged diaspora (stock)	-0.0057	-0.0133
	(0.0085)	(0.0086)
GDP per capita	0.0290***	0.0276***
	(0.0044)	(0.0044)
Population	-0.0015***	-0.0000
	(0.0004)	(0.0000)
Lagged skill ratio	0.1636***	0.1145**
	(0.0569)	(0.0607)
Sargan P-Value	0.1168	0.3614
AR(1)	0.0001	0.0006
AR(2)	0.4753	0.9868
Countries	39	39
Observations	262	262

* p<0.1; ** p<0.05; *** p<0.01

5 Conclusion

Occupation-skill mismatch is a well known labour market phenomenon especially for immigrants. The implications of this mismatch especially on wages have been extensively studied. However, little is known about the impacts of mismatch on selection of immigrants. In this paper, using panel data of immigrants from 40 countries in Portugal, we show that indeed occupation-skill mismatch affects selection of immigrants. The results suggest that the probability of overeducation leads to negative selection of immigrants. Furthermore, we show that when probability of correct matches increases, the high skilled are more likely to migrate than the low skilled (positive selection).

The results from this paper have implications for sorting of immigrants. From the results, we observe that due to overeducation, the high skilled are less likely to migrate compared to the low skilled. This finding raises another interesting question for further research. How does occupation-skill mismatch affects sorting of immigrants among different destination countries? That is, will migrants prefer migrating to countries with better skill matches. The framework developed in this paper can be extended to include more than one destination country. To answer this question, one would need data on not only one country as we have in this paper, but also data on immigrants in other destination countries.

From a policy perspective, these findings also have some important implications.

Receiving countries' selective policies aimed at attracting high skilled immigrants should also inculcate reducing occupation-skill mismatch probably through degree recognition and standardization in collaboration with sending countries.

References

Alba-Ramírez, Alfonso and Maria Jesús San Segundo (1995). The returns to education in Spain, *Economics of Education Review* 14.2, pp. 155–166.

Ambrosini, J. William and Giovanni Peri (2012). The Determinants and the Selection of Mexico - US Migrants, *The World Economy* 35.2, pp. 111–151.

Arellano, Manuel and Stephen Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *The review of economic studies* 58.2, pp. 277–297.

Batista, Catia, Aitor Lacuesta, and Pedro C. Vicente (2012). Testing the brain gain hypothesis: Micro evidence from Cape Verde, *Journal of Development Economics* 97.1, pp. 32–45.

Batista, Catia, Tara McIndoe-Calder, and Pedro C Vicente (2017). Return Migration, Self-selection and Entrepreneurship, *Oxford Bulletin of Economics and Statistics* 79.5, pp. 797–821.

Batista, Catia and David McKenzie (2018). "Testing Classic Theories of Migration in the Lab". *Mimeo*, Universidade Nova de Lisboa and World Bank.

Batista, Catia, Julia Seither, and Pedro C Vicente (2016). *Migration, Political Institutions, and Social Networks in Mozambique*. Tech. rep. NOVAFRICA Working Paper.

Batista, Catia and Pedro C. Vicente (2011). Do Migrants Improve Governance at Home? Evidence from a Voting Experiment, *The World Bank Economic Review* 25.1, pp. 77–104.

Beine, Michel, Frédéric Docquier, and Çağlar Özden (2011). Diasporas, *Journal of Development Economics* 95.1, pp. 30–41.

Beine, Michel, Frédéric Docquier, and Hillel Rapoport (2001). Brain drain and economic growth: theory and evidence, *Journal of Development Economics* 64.1, pp. 275–289.

— (2008). Brain drain and human capital formation in developing countries: winners and losers, *The Economic Journal* 118.528, pp. 631–652.

Bhagwati, Jagdish and Koichi Hamada (1974). The brain drain, international integration of markets for professionals and unemployment, *Journal of Development Economics* 1.1, pp. 19–42.

Blundell, Richard and Stephen Bond (1998). Initial conditions and moment restrictions in dynamic panel data models, *Journal of econometrics* 87.1, pp. 115–143.

Borjas, George J. (1987). Self-Selection and the Earnings of Immigrants, *The American Economic Review* 77.4, pp. 531–553.

Cabral, Sonia and Cláudia Duarte (2013). *Mind the gap! The relative wages of immigrants in the Portuguese labour market*. Tech. rep. Banco de Portugal, Economics and Research Department.

Carneiro, Anabela, Natércia Fortuna, and José Varejão (2012). Immigrants at new destinations: how they fare and why, *Journal of Population Economics* 25.3, pp. 1165–1185.

Chiquiar, Daniel and H. Gordon Hanson (2005). International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States, *Journal of Political Economy* 113.2, pp. 239–281.

Chiswick, Barry R and Paul W Miller (2009). The international transferability of immigrants' human capital, *Economics of Education Review* 28.2, pp. 162–169.

Duncan, Greg J and Saul D Hoffman (1981). The incidence and wage effects of overeducation, *Economics of Education Review* 1.1, pp. 75–86.

Grogger, Jeffrey and Gordon H Hanson (2011). Income maximization and the selection and sorting of international migrants, *Journal of Development Economics* 95.1, pp. 42–57.

Grubel, Herbert B. and Anthony D. Scott (1966). The International Flow of Human Capital, *The American Economic Review* 56.1/2, pp. 268–274.

Haque, Nadeem U. and Se-Jik Kim (1995). "Human Capital Flight": Impact of Migration on Income and Growth, *Staff Papers (International Monetary Fund)* 42.3, pp. 577–607.

Hartog, Joop (2000). Over-education and earnings: where are we, where should we go?, *Economics of Education Review* 19.2, pp. 131–147.

Kiker, BF and Maria C Santos (1991). Human capital and earnings in Portugal, *Economics of Education Review* 10.3, pp. 187–203.

Kiker, Billy Frazier, Maria C Santos, and M Mendes De Oliveira (1997). Overeducation and undereducation: Evidence for Portugal, *Economics of Education Review* 16.2, pp. 111–125.

Mattoo, Aaditya, Cristina Neagu Ileana, and Çağlar Özden (2008). Brain waste? Educated immigrants in the {US} labor market, *Journal of Development Economics* 87.2, pp. 255–269.

McFadden, Daniel L (1984). Econometric analysis of qualitative response models, *Handbook of econometrics* 2, pp. 1395–1457.

McKenzie, David and Hillel Rapoport (2010). Self-selection patterns in Mexico-US migration: the role of migration networks, *The Review of Economics and Statistics* 92.4, pp. 811–821.

Moraga, Jesus Fernandez-Huertas (2011). New evidence on emigrant selection, *The Review of Economics and Statistics* 93.1, pp. 72–96.

Orrenius, Pia and Madeline Zavodny (2005). Self-selection among undocumented immigrants from Mexico, *Journal of Development Economics* 78.1, pp. 215–240.

Pires, Armando J Garcia (2015). Brain drain and brain waste, *Journal of Economic Development* 40.1, pp. 1–34.

Quinn, Michael A. and Stephen Rubb (2005). The Importance of Education-Occupation Matching in Migration Decisions, *Demography* 42.1, pp. 153–167.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings, *Oxford Economic Papers* 3.2, pp. 135–146.

Schiff, Maurice (2005). *Brain gain: claims about its size and impact on welfare and growth are greatly exaggerated* vol. 1599. World Bank Publications.

Sjaastad, Larry A. (1962). The Costs and Returns of Human Migration, *Journal of Political Economy* 70.5, pp. 80–93.

Tsai, Yuping (2010). Returns to overeducation: A longitudinal analysis of the US labor market, *Economics of Education Review* 29.4, pp. 606–617.

United Nations, Department of Economic & Social Affairs Population Division (2016). *International Migration Report 2015: Highlights ({ST/ESA/SER.A/375})*.

Verdugo, Richard R. and Naomi Turner Verdugo (1989). The Impact of Surplus Schooling on Earnings: Some Additional Findings, *The Journal of Human Resources* 24.4, pp. 629–643.

Yang, Dean (2008). International Migration, Remittances and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks, *The Economic Journal* 118.528, pp. 591–630.

6 Appendix

6.1 Definition of Variables

	Multinomial Logit Model
Dependent Variables	Description
$Ln(OE_{it}/M_{it})$	log ratio of overeducation to matched of individual i in time t
$Ln(UE_{it}/M_{it})$	log ratio of undereducation to matched of individual i in time t
Independent Variables	Description
Experience	Age - 6 - years of education
Tenure	Current year - year the worker started working for the firm
Female	Dummy variable equal 1 for female and equal 0 for male
High Skilled	Dummy variable equal for workers with more than 12 years of education and zero otherwise
Years since migration	This is a proxy for year the worker entered the country. Following Cabral and Duarte (2013), we proxy the year of migration with the minimum of the year the worker first appeared in the database. This is achieved my tracing back the worker ID number to the year that it appears from the data base starting from 1986.
	Selection Equation
Dependent Variable	Description
Ln(Skill Ratio)	log ratio of the flow of high skilled to low skilled immigrants.
Independent Variables	Description
Overeducation differences	Probability of overeducation for high skilled migrant - probability of overeducation for low skilled migrant.
Occupation-skill match differences	Probability of occupation-skill match for high skilled migrant - probability of occupation-skill match for low skilled migrant.
Wage differences	Average high skilled hourly wage - Average low skilled hourly wage.
Lagged Diaspora	Stock of immigrants from the origin in the previous year.
GDP per capita	Gross domestic product per capita of the country of origin.
Population	Total population of the country of origin.
Lagged skill ratio	log ratio of the flow of high skilled to low skilled immigrants in the previous.

6.2 Stock of Workers

Year	Natives	Migrants
2002	2,543,691	95,792
2003	2,552,969	116,897
2004	2,604,026	125,956
2005	2,920,259	146,620
2006	2,859,951	146,190
2007	2,952,052	153,185
2008	2,005,729	168,552
2009	2,889,639	155,139