Haylee Ham
MACS 30200

Global International Migration Trends: Characteristics of Origin and Destination Countries that Motivate Migration

I. Data

Until recently, global bilateral migration data has been sparse and therefore studies at a global scale have been limited. This paper will use the recently published dataset from the Wittgenstein Centre for Demography and Global Human Capital. This dataset contains global bilateral flows between 196 countries from 1990 to 2010 in five year increments. The authors of the dataset note that "migration flow data is often incomplete and not comparable across nations", however, using migrant stock tables, Abel and Sander have created a robust and novel dataset that is comparable across countries and accurately captures origins and destinations for migrants who have permanently changed their country of residence over five year periods. This dataset has been created to study migration trends between all pairs of countries, wherein each country is both the origin and destination country for every other country in a row of the data. Each row is thus a pair of countries, one destination and one origin, and the number of migrants who permanently changed their residence over a five-year period from the origin to the destination country in that row. This was accomplished by linking changes in migrant stock data provided by the United Nations over time. Missing data methods were used to estimate the migrant flows necessary to meet the differences observed in the migrant stock data (Abel and Sander, 2014).

In order to observe the preferences of the migrant population, I will make use of the dataset published by Abel and Sander (2014) and add country-level development and economic indicators. These indicators have been taken from World Bank's world development indicators dataset. The bilateral migration flows data set can be accessed on the following website: http://www.global-migration.info/. The world development indicators can be accessed at the World Bank's website: http://data.worldbank.org/data-catalog/world-development-indicators.

In its raw form, the bilateral migration flows dataset has a variable for migration flows between all possible pairs of countries for each five year period. For the purposes of this study, which is concerned with general trends of migration rather than time-series effects, I have reshaped the data so that all migration flow counts occur within one column and a separate column to indicate the five year interval has been appended. The world development indicators are observed yearly and have been averaged into the same five year increments present in the bilateral migration flows dataset. The averages for all countries have then been matched to the

bilateral migration flows dataset so that each migration flow observation has both origin and destination country characteristics associated with it. This results in a dataset with 103,684 observations of bilateral migration flows and 7 world development indicators for each observation.

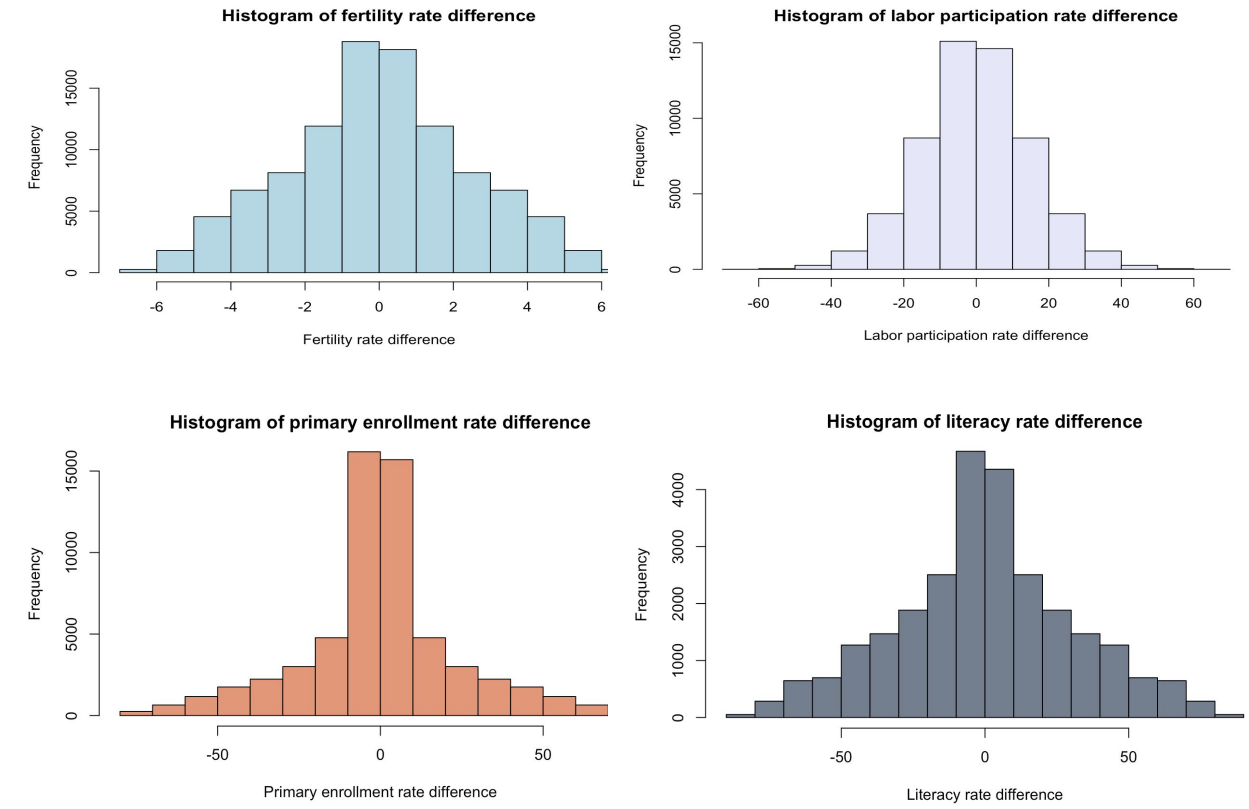
The world development indicators included in this dataset are fertility rate, literacy rate, per worker GDP, labor participation rate, safety net coverage rate, working age rate, and primary enrollment rate. Fertility rate is measured as the number of children that a woman would birth if she were to live until the end of her childbearing years and birth the number of children indicated by age-specific fertility rates for the time period. These rates are developed by using registered live births or census data. Literacy rate is the number of individuals who are 15 years old and above who are capable of both reading and writing short simple statements. Per worker GDP is the GDP for a country divided by total employment and measured in constant 2011 purchasing power parity (PPP) international dollars, which has the same PPP as U.S. dollars in the United States. Labor participation rate is the percentage of the total population ages 15 and above that is economically active, including those who are employed and those who are unemployed for a temporary reason. Safety net coverage rate is the percentage of the population who are benefitting from cash transfers and other social programs such as disability benefits, school feeding, food stamps, housing allowances, and health subsidies. Working age rate is a measure of the percentage of the population that is between 15 and 64 years old. This age range is generally considered to be of working age. Primary enrollment rate is the percentage of primary school age children enrolled in school. The variables of interest in the resulting dataset are summarized in the table below.

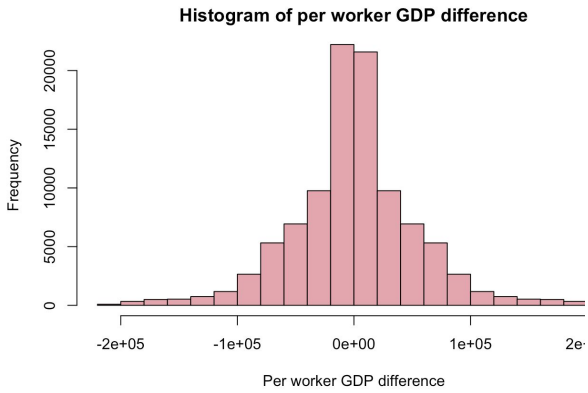
Table 1: Summary Statistics for Key Variables

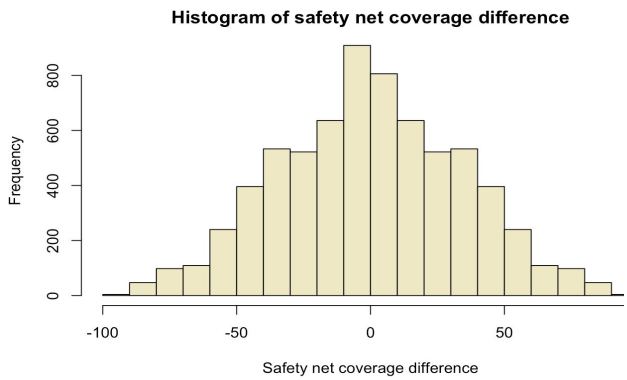
Variable	Min	Median	Mean	Max
Migration flow (count)	0	0	1191	2,677,763
Fertility rate (count)	1.15	2.83	3.37	7.75
Labor participation rate (%)	27	59	60	93
Per worker GDP ($)	508	20,200	34,708	211,619
Literacy rate (%)	11	85	76	100
Primary enrollment rate (%)	21	94	86	100
Safety net coverage (%)	0	33	32	91
Working age rate (%)	47.9	61.7	60.7	83.8

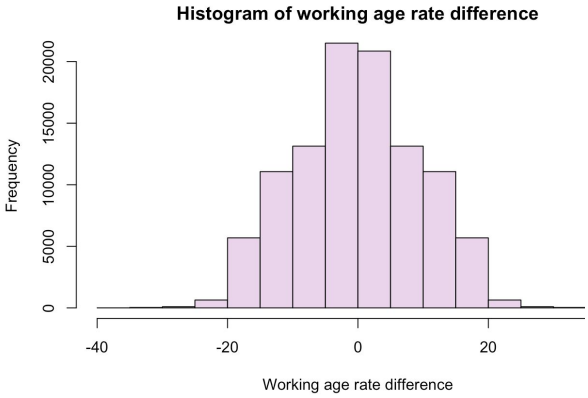
For the purposes of this study, a simple difference of the values of these development indicators between destination and origin country will be taken for each migration flow observation. These differences will allow the net effect of each of these indicators to be observed between all possible country pairings. Once the difference is performed, the development indicators are distributed symmetrically, with some distributions approximating the normal distribution and others showing higher peaks and heavier tails.

Figure 1: Histograms of World Development Indicator Differences









II. Methodology

In order to explain the variation in migration flows between countries, I will first employ a linear model with count of migrants between each pairing of countries as the dependent variable. As independent variables, the model will contain differences of the world development indicators between the destination and origin country of each migration flow. Specific world development indicators were chosen as representations of the theoretical constructs presented in the literature review. Specifically, per worker GDP and labor participation rate are used to measure the neoclassical theory of migration introduced by Harris and Todaro (1970), which states that possible income in a country and the probability of securing employment in that country are large factors when choosing a destination. Fertility rate and working age rate are the variables used in this study to approximate the effect of the demographic shift theory posited by Hugo (1998), which states that destination countries have lower fertility rates and an aging population compared to origin countries. Borjas' "welfare magnate effect" (1987) will be tested using safety net coverage rate as a proxy for the generosity of a nation's welfare program. Primary school enrollment rate will be used as a measure of the current strength of the education system in a country, an amenity which families consider very important when considering migrating (Adepoju, 2000) and literacy rate will measure the proportion of individuals who have

obtained at least a basic education, a necessary skill set for someone desiring to migrate (Castles, 2002).

The neoclassical theory and the demographic shift theory are the most supported in the literature. Because of this, the model for this study has been built around those two theories:

$$migration\_flow_{i,j} = \beta_0 + \beta_1\, per\_worker\_GDP_{i,j} + \beta_2\, labor\_participation\_rate_{i,j} + \beta_3\, per\_worker\_GDP_{i,j}{:}labor\_participation\_rate_{i,j} + \beta_4 fertility\_rate_{i,j} + \beta_5\, working\_age\_rate_{i,j} + \beta_6 fertility\_rate_{i,j}{:}working\_age\_rate_{i,j} + \beta_7\, primary\_enrollment\_rate_{i,j} + country\_fixed\_effects_i + time\_fixed\_effects_t + \varepsilon$$

Each of the independent variables are the differences between the origin and destination countries Since the neoclassical theory and the demographic shift theory both hypothesize that the effects of one indicator will fluctuate depending on the value of another indicator, interaction terms were included for per worker GDP and labor participation rate and also fertility rate and working age rate. Primary enrollment rate was also included as the literature has considered the strength of a country's education system to be a strong indicator for the ability a migrant has to relocate. The model also controls for all 196 origin country fixed effects and all four time periods fixed effects using dummy variables. Since the independent variables are measures of the difference between the origin and destination country, controlling for fixed effects in the origin country will not control for the variation present in the differences derived from the indicator variables.

As an initial step, I ran factor regressions in order to test the individual effects each of the indicators has on migration flows. I performed factor regressions on both the indicators included in the model and indicators that were not included in the model but that previous studies have considered important. In these regressions, the migration flow was the dependent variable and each indicator became the independent variable. In order to isolate the effect of the indicator, I also controlled for country and time fixed effects. The results of the factor regressions are presented in Table 2 below.

Table 2: Factor Regressions of World Development Indicators

Variable	Estimate	Std. Error	p-value
Fertility rate	-242.4	35.4	7.47e-12*
Labor participation rate	30.9	8.28	1.88e-04*
Per worker GDP	2.61e-02	1.69e-03	5.53e-54*
Literacy rate	7.6	4.09	6.30e-02
Primary enrollment rate	33.37	5.03	3.36e-11*
Safety net coverage	7.12	6.4	2.66e-01
Working age rate	90.05	9.14	6.69e-23*

** Significant at a 0.05 level*

From the table above, it can be seen that fertility rate, labor participation rate, per worker GDP, primary enrollment rate, and working age rate are all significant at at least a 0.05 level. Literacy rate and safety net coverage do not appear to be significant factors in explaining migration trends, even though previous literature has theorized the importance of these factors (Castles, 2002 and Borjas, 1987).

Because the interaction terms for the neoclassical and demographic shift theories are vital to this model, I also created factor regressions for the interaction terms between per worker GDP and labor participation rate and also fertility rate and working age rate. The results for those factor regressions can be seen in Table 3 below.

Table 3: Factor Regressions for Interaction Terms of World Development Indicators

Variable	Estimate	Std. Error	p-value
Per worker GDP	3.73e-02	2.46e-03	5.76e-52*
Labor participation rate	3.39e+01	8.36e+00	4.95e-05*
Per worker GDP:Labor participation rate	2.68e-04	1.24e-04	3.03e-02*
Fertility rate	534.6	85.45	3.96e-10*
Working age rate	204.6	21.77	5.50e-21*
Fertility rate:Working age rate	18.3	2.56	9.05e-13*

** Significant at a 0.05 level*

Both of these interaction terms produced significant results when regressed on the migration flows and controlling for origin country and time fixed effects.

As an initial test of the model, I have used the ordinary least squared method of linear regression. The results of the OLS model can be seen in Table 4 below.

Table 4: OLS Regression of Model

Variable	Estimate	Std. Error	p-value
Per worker GDP	3.70e-02	4.23e-03	2.48e-18*
Labor participation rate	4.11e+01	1.22e+01	7.44e-04*
Fertility rate	-1.34e+02	2.44e+02	5.83e-01
Working age rate	3.42e+01	5.58e+01	5.40e-01
Primary enrollment rate	3.03	1.37e+01	8.25e-01
Per worker GDP: labor participation rate	2.64e-04	1.72e-04	1.24e-01
Fertility rate: working age rate	3.90e-01	7.32	9.58e-01

** Significant at a 0.05 level*

Per worker GDP is significant in this model at a 95% confidence level. The estimate states that as the difference between per worker GDP in the origin and destination country increases, the migration flow to the destination country (with a higher per worker GDP) will increase. Labor participation rate also appears to be a significant indicator. As the gap between labor participation rate in the origin and destination country increases, migration flows will increase to the destination country due to the higher chance of finding employment in that country. However, the interaction effect of these two variables is not significant at a 95% level. Similarly, fertility rate and working age rate and their interaction term are not significant when placed in the model with the neoclassical indicators of per worker GDP and labor participation rate. The primary enrollment rate (an indicator of the strength of the education system in the destination country as compared to the origin country) is also not a significant predictor of migration flows.

References

Abel, Guy J., and Nikola Sander. “Quantifying Global International Migration Flows.” *Science*, vol. 343, no. 6178, 2014, pp. 1520-1522., http://science.sciencemag.org/content/343/6178/1520.

Adepoju, A. "Issues and Recent Trends in International Migration in Sub-Saharan Africa." *International Social Science Journal*, vol. 52, no. 165, 2000, pp. 383-394*, SCOPUS*, www.scopus.com.

Castles, S. "Migration and Community Formation Under Conditions of Globalization." *International Migration Review*, vol. 36, no. 4, 2002, pp. 1143-1168*, SCOPUS*, www.scopus.com.

Harris, John R., and Michael P. Todaro. “Migration, Unemployment and Development: A Two-Sector Analysis.” *The American Economic Review*, vol. 60, no. 1, 1970, pp. 126–142., www.jstor.org/stable/1807860.

Hugo, G. "The Demographic Underpinnings of Current and Future International Migration in Asia." *Asian and Pacific Migration Journal*, vol. 7, no. 1, 1998, pp. 1-25*, SCOPUS*, www.scopus.com.