# The GMig2 Data Base:
# Extending the GTAP 8 to Include
# Global Bilateral Migration, Wages and Remittances[1]

**Terrie L. Walmsley**[2]
**Angel Aguiar**[3]
**Zekarias Hussein**[4]
**and**
**Christopher R. Parsons**[5]

February 2013

[2] Terrie Walmsley is Associate Professor and Director of the Center for Global Trade Analysis, Purdue University, 403 W. State St, West Lafayette IN 47907. Ph: +1 765 494 5837. Fax: +1 765 496 1224. Email: twalmsle@purdue.edu.
[3] Angel Aguiar is a Research Fellow the Center for Global Trade Analysis, Purdue University.
[4] Zekarias Hussein is a PhD Student at the Center for Global Trade Analysis, Purdue University.
[5] Christopher Parsons is a Research Officer at the International Migration Institute, University of Oxford.

# 1.  Introduction

This document describes the construction of the Global Bilateral Migration Data Base, the GMig2 Data Base, based on the GTAP 8 Data Base.  To avoid confusion we will refer to this database as the GMig2 8_0 Data Base where 8 refers to the version of the underlying GTAP Data Base and 0 refers to the first or original version of the migration data.[6]

Recent developments by Özden, Parsons, Schiff and Walmsley (2011) to construct a time series of bilateral matrix of foreign population have allowed us to update the previous GMig2 7_0 database of labor, remittances, and wages.  The purpose of this document is to outline how this data is combined with other data on wages and remittances to create a database which can be used to model the impact of labor movements.

The GMig2 8_0 Data Base is based on and consistent with the GTAP 8 Data Base.  The GTAP 8 Data Base contains input-output data on 129 regions and 57 commodities, as well as detailed bilateral trade, transport and protection information (Narayanan et al., 2012).  In addition to the GTAP Data Base and the Özden, Parsons, Schiff and Walmsley (2011) we also obtained remittance data from the IMF; participation rates obtained from the ILO LABORSTA database website (ILO, 2012); skill splits estimated from data obtained from LABORSTA and Docquier et al. (2010).  The migration labor force data and total remittances are constructed for 244 countries and then aggregated up to the GTAP countries, where wages and incomes can then be determined.  In this way the migration database can be updated as new countries are incorporated into the GTAP Data Base.

The document is divided into five sections.  Following the introduction, section two outlines the data sources used to derive the Global Bilateral Migration Data Base.  Section three explains the procedures used to obtain migrant labor by skill.  In section four, we examine the techniques used to determine wages and remittances.  Section five then concludes the paper.


# 2.    Sources of Data

The following data are utilized in the construction of the Global Bilateral Labor Migration (GMig2) Data Base:

   a.  Labor income data is obtained from the GTAP 8 Data Base.  The GTAP Data Base covers 129 regions, 5 endowments (skilled and unskilled labor, capital, natural resources and land) and 57 sectors.
   b.  The number of foreigners by home and host countries is obtained from Özden, Parsons, Schiff and Walmsley (2011).[7]  This is a matrix of 226 by 226 countries, which was further extended to

---

[6] Therefore, GMig2 8.1_0 Data Base would be the original version (_0) of the Gmig2 Data Base, based on version 8.1 of the GTAP Data Base.  Should updates be made to the migration data then the number after the underscore would adjust.

[7] No additional migration is collected; this is simply done to be consistent with GTAP, which moved to a standard 244 country list.  The additional 18 countries are usually small protectorates; in fact some do not even have an estimated population (e.g., Antarctica).

244 by 244.[8] Özden, Parsons, Schiff and Walmsley (2011) collected data on both the foreign born and nationality of residents over time, primarily from census data. The resulting foreign born data, filled using the methods described in Özden, Parsons, Schiff and Walmsley (2011), are utilized. Note that this data are based on foreign born and no account is taken of the length of stay, hence the data include both permanent and temporary migrant labor in the host region by home country at a given point in time.

c. Total remittances received are obtained from the IMF Balance of Payments Statistics and include both remittances and workers compensation. Data is available for 132 countries.

d. Participation rates for 188 countries were obtained from LABORSTA (ILO, 2012).

e. Data on the split between skilled and unskilled were obtained from two different sources: a) Docquier et al. (2010) provide skilled-unskilled labor split data for migrant workers in 195 host and home regions. This is an updated version of the database constructed by Docquier and Markouk (2004); and b) data on the skill level of labor for 142 countries was also obtained from LABORSTA (ILO, 2012).

f. Purchasing power parity data for 182 countries was obtained from the World Bank's World Development Indicators dataset. This was then extended to 244 countries using regional averages.

# 3. The Global Migrant Labor Force

In this section we use the bilateral foreign population stock data (Özden, Parsons, Schiff and Walmsley, 2011), along with some additional data and assumptions, to obtain bilateral foreign labor forces by skill.

## 3.1 The Global Migrant Population

The bilateral foreign population database constructed by Özden, Parsons, Schiff and Walmsley (2011) underlies all of the new migration data in the GMig2 8_0 Data Base.

## 3.2 Participation Rates

The Özden, Parsons, Schiff and Walmsley (2011) database is based on foreign born and does not distinguish between those in the labor force and those who are not. In order to convert the population data to labor forces, additional data on the participation rates are collected from LABORSTA (ILO, 2012). It is then assumed that the participation rates of migrant labor are the same as the participation rates in the home region. This means that migrant labor are assumed to move with their families. This is likely to be an overestimate of migrant workers because many may have migrated when they were young and have had their children in the host country; also some migrants will be temporary and hence have not brought their families. Unfortunately, there are no data on the extent to which workers migrate with their families or not. Figure 1 shows the participation rates for selected regions. Participation rates are generally between 40 and 60 percent.

---

[8] The number of countries was increased because the GTAP 8 Data Base is based on data for a standard list of 244 countries. The extension was made using the census data and the methods outlined in Parsons et al. (2011).

## *3.3   Skill Levels*

The skill level of migrants is an important issue for policy makers in both the labor exporting and labor importing economies.  In labor importing economies skilled migrant labor are generally more easily accepted in their host economies, while unskilled labor still raise significant concerns despite potential gains. In the labor exporting economies, on the other hand, the loss of skilled labor is often a cause for concern, for example, the migration of doctors to the United Kingdom from the Philippines or Africa. In general migrant labor are thought to be more skilled than the workers they leave behind, particularly those moving to developed economies like the United States.  In order to determine the skill levels of migrant labor we draw heavily on the education database developed by Docquier et al. (2010).

**Figure1. Participation Rates by Aggregate Regions**
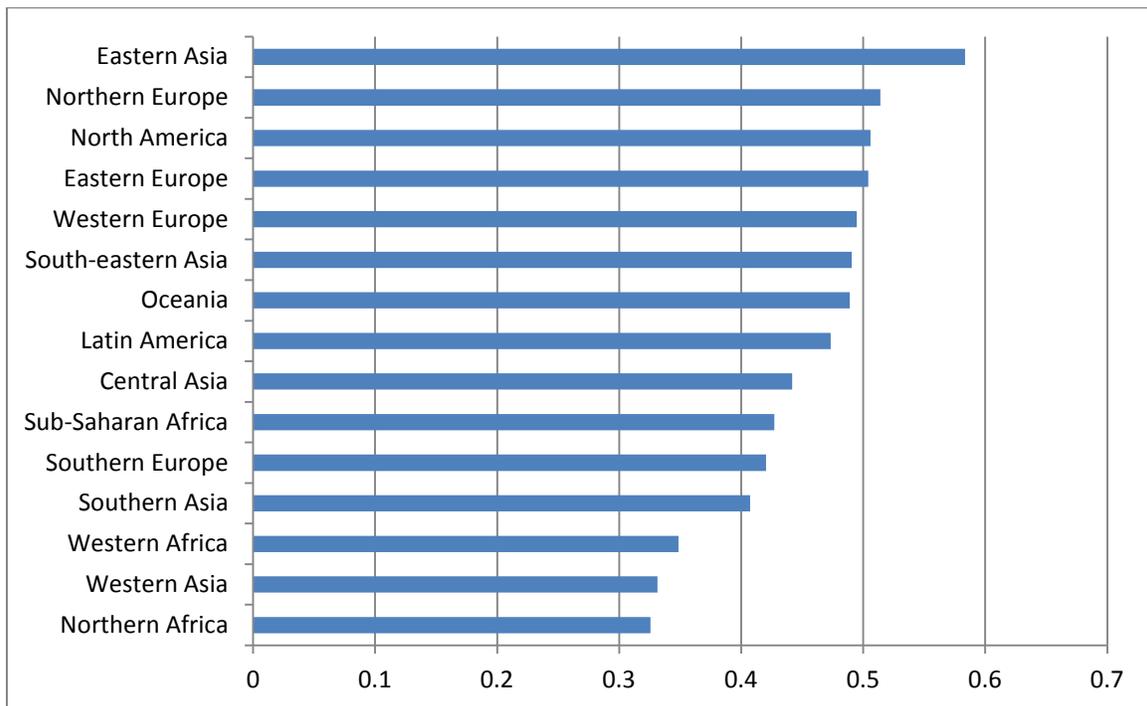


In the GTAP Data Base the skill splits are defined in terms of occupation (Liu et al., 1998). Unfortunately migration data by occupation are scarce and we have had to rely on education data. Determining the skill splits of labor by home and host regions proceeded as follows:

1. Initially data were collected on the numbers of skilled and unskilled workers by occupation in each country from LABORSTA (ILO, 2012).  Unfortunately, shares could not be obtained for all 244 countries; and hence shares for the existing economies were used as initial estimates for the remaining countries.  These were then combined with the wages data in the GTAP Data Base to obtain the relative wages of skilled and unskilled workers in a given region.

2. Once the share of skilled labor was obtained for the region/country as a whole, we had to obtain the skill characteristics of the migrant labor.  A number of sources and assumptions were investigated:

    a. First foreign labor was assumed to have the same skill characteristics as their home region.  This assumption ensures that those countries with more skilled labor supply more

4

skilled migrant labor, e.g., more skilled labor is supplied by developed economies such as the USA, Europe and Australia as a share of total migrant labor. This method, however, failed to pick up some of the more interesting features we might expect, e.g., the tendency for India and China to export skilled labor, despite the fact that both have very low shares of skilled labor at home. Mattoo, Neagu and Özden (2005) investigate the skill level of migrant labor by education in the USA and find that the skill levels of migrant labor do not necessarily follow the skill levels of the home residents.

b.  Docquier et al. (2010) completed a database on educational attainment of migrant labor in 195 host regions from 195 home regions. The benefit of this database was that it did pick up some of the features referred to in (a) above; however, data was based on education, rather than occupation. Harrison et al. (2003) also use education to examine the skill levels of migrant labor. Despite the inconsistency between these data and the GTAP Data Base, we chose to use the education data from Docquier et al. (2010), because we feel that it does allow us to capture these regional patterns in the skill levels of migrants.

3.  Once again, skill shares could not be obtained for all migrant labor in all countries, so the missing shares were filled using an average skill share overseas, with the method of calculation depending on the type of missing data. Specifically, there were three cases where skill share data was unavailable.

a.  The first case was where there was data on the skill of migrant labor from region r, but not distinguished by all of the host regions. Because the Docquier et al. (2010) database covers 195 of our 244 host countries, one option is to use the average skill share for the migrant labor from r, across all host regions for which data was available, to fill in the missing values. In this case migrant workers who leave region r have the same characteristics as other migrant labor that leaves region r, regardless of their destination. For example if India sent skilled workers to the USA, but unskilled migrants to the Middle East, using the average would underestimate the skill of Indian migrants in the USA and overestimate those in the Middle East. This method has the limitation that there may be considerable differences in the skill levels of migrant labor by destination, as well as source. Ideally we would like to be able to gather data for each country or region but this is not available. Hence the average skill share for the migrant labor from r, across all host regions for which data was available, is used to fill in the missing values.

b.  The second case was where there was data for migrant labor in the host region c, but not distinguished by all home regions. In a manner similar to that used above, the missing skill shares for a host region c were taken to be the average skill share of all migrant labor in that region. Since there was no data on the characteristics of the average migrant from region r, it was assumed that migrant labor from region r, destined for region c would have similar characteristics to other migrant labor in region c. Hence if migrant labor to the USA tended to be skilled, then migrant labor from the missing home regions tended to be more skilled. Since the Docquier et al. (2010) database is quite comprehensive, this was not a major issue and had little impact on the resulting skill shares when aggregated to the GTAP Data Base's 129 regions.

c.  The third case was where there was no data on the skill shares for any migrant labor from region r or located in region c (where r is the home country and c is the host country). These gaps were filled by using the average skill share from the Docquier et al. (2010) database.

The resulting skill shares confirm that skilled workers are more mobile than unskilled workers. In most cases the host country's share of skilled imported labor is higher than the country's own share of skilled labor. Figure 2 shows the skill shares of permanent residents and migrant labor.

**Figure 2: Skill Shares of permanent residents and migrant labor exported and imported**
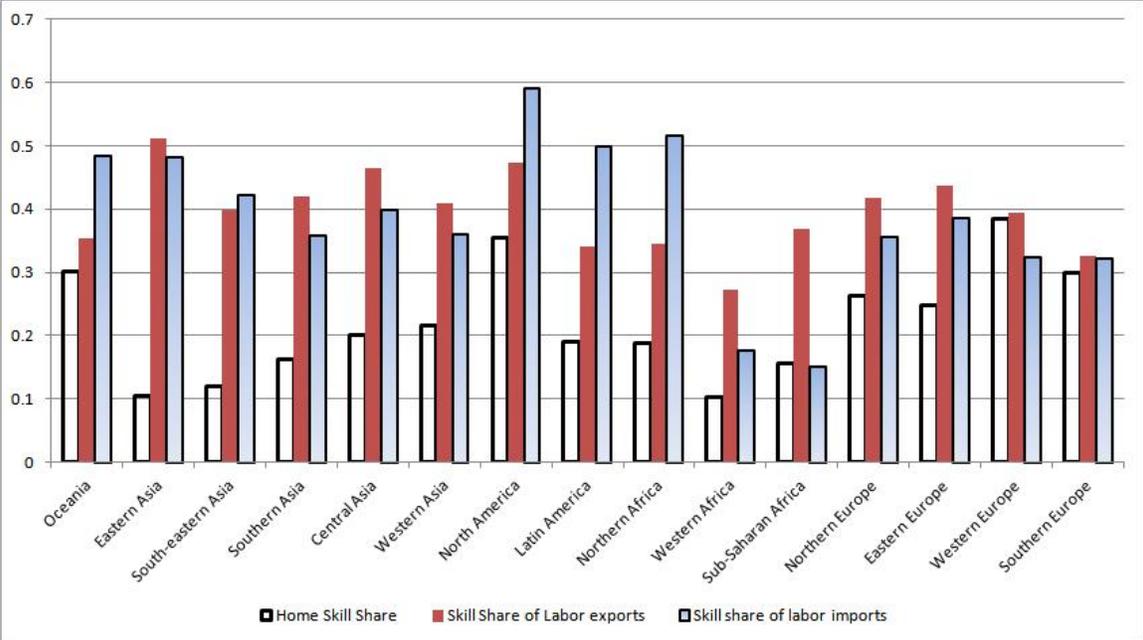
# 4.    *Wages and Remittances*

## 4.1    *Wages and Labor income*

Labor income earned by migrant labor is required in order to examine the impact of policies on migrant incomes. No data is available on either labor income or wage rates earned by migrant labor on a global basis and hence must be derived. Labor income earned within a region by all workers is obtained from the GTAP 8 Data Base.

Wage rates of workers of skill i, from region r, located in region c ($W_{i,r,c}$) are assumed to equal the home wage ($HW_{i,r}$) in region r, plus a proportion (BETA) of the difference between the host and home wage ($HW_{i,c}$ - $HW_{i,r}$):

$$W_{i,r,c} = HW_{i,r} + BETA_{i,r,c} \; x \; (HW_{i,c} - HW_{i,r}) \qquad\qquad (1)$$

where:   BETA is the proportion of the difference obtained by a person of labor type i migrating from region r to region c.

This equation stems from the fact that the wages of migrant labor are generally lower than the wages prevailing in the host country (Borjas, 2000). The extent to which wages are lower is determined by BETA. BETA is set equal to 0.75 when the home wage is less than the host country wage; e.g., when a migrant moves from a developing to a developed economy. The choice of a high value for BETA (0.75) reflects the fact that the workers are more permanent and therefore earn a larger proportion of the host countries wage and productivity. BETA is set to 0.3 when the host wage is less than the home wage; e.g., when a person moves from a developed to a developing country wages are not likely to decline significantly. The catch-up parameter (BETA) is obviously crude, but in the absence of information we do not have a better estimate. Borjas (2000) reports eventual catch-up of over 100% for permanent migrant labor (i.e., overtaking local wages), but for temporary workers the catch-up will inevitably be significantly smaller. Ideally we would need to know how long migrants had been away from home to do this properly, however such data is not available globally.

The labor income earned by all permanent residents and migrant workers must also equal the total returns to labor from the GTAP Data Base in order to be consistent with the GTAP Data Base, hence the last stage is to adjust the wage rates to ensure balance.

Figure 3 depicts the average (nominal) wages of permanent residents in selected countries. As expected the wages of skilled are greater than those of unskilled workers, and wages in developed countries are higher than those in developing countries.

**Figure 3: Average Wages of Permanent Residents by Aggregate Regions**
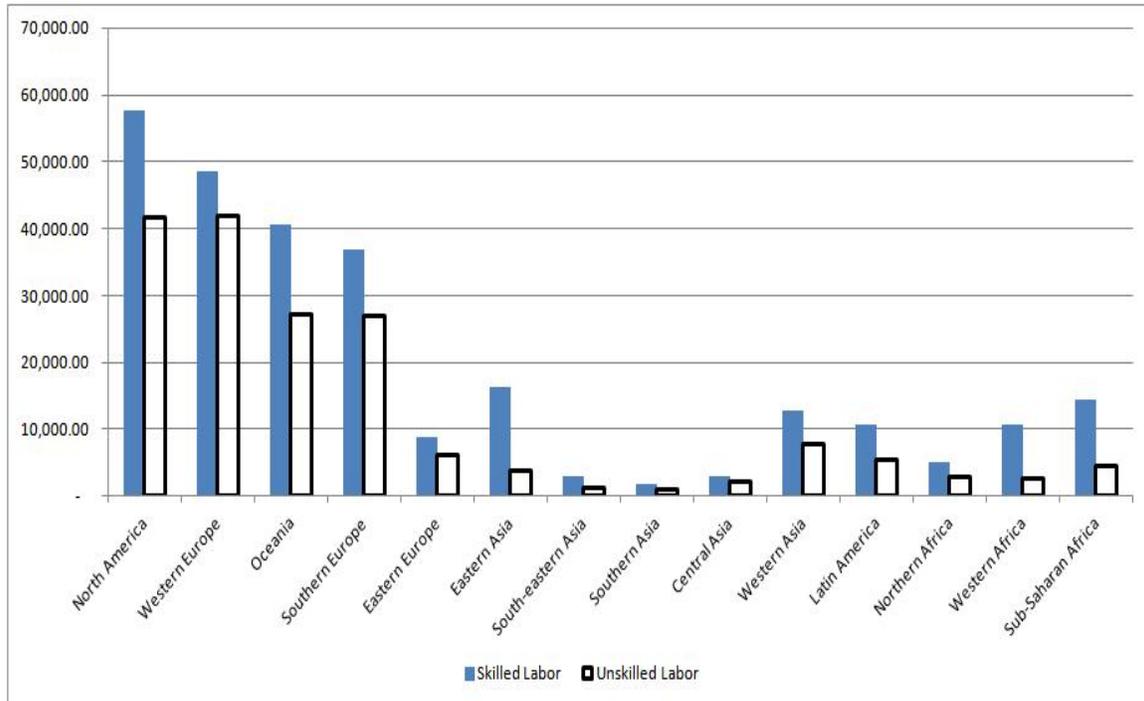


## *4.2 Purchasing Power Parity and Real Wages*

We are also interested in the impact of movements of labor across countries therefore, it is also important to take into consideration the differences in purchasing power between countries (Figure 4). For this reason we also obtain measures of purchasing power parity from the World Bank. The resulting real wages of permanent residents are shown in Figure 5.

## *4.3 Remittances*

Key in any analysis of the gains/losses from labor migration is the extent to which remittances received by home economies compensate the home country for outward migration. Furthermore remittances have grown considerably in recent years and often exceed official aid flows and, in the case of Africa, foreign direct investment (OSSA-UN, 2005).

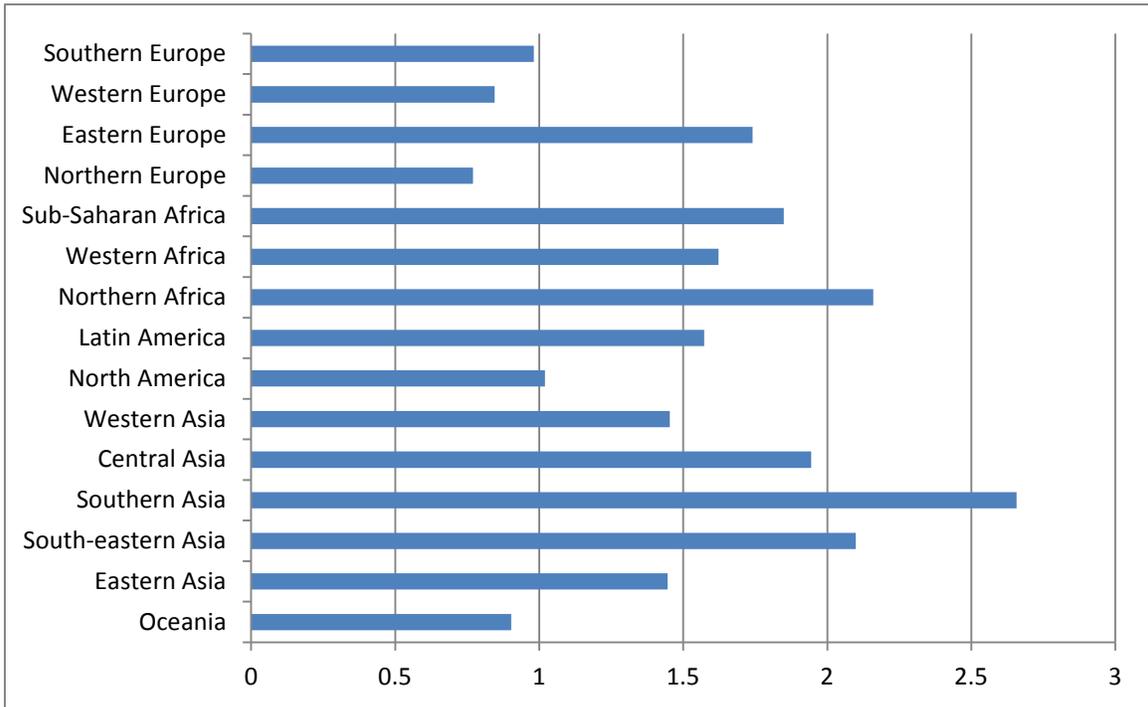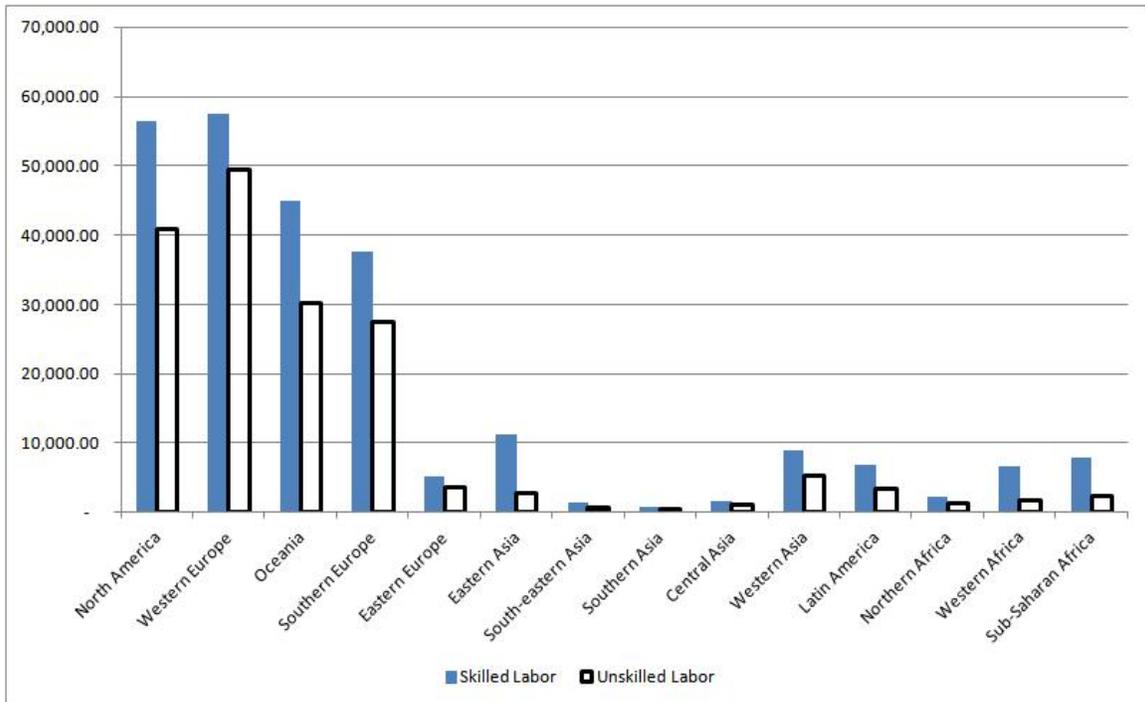**Figure 4: Purchasing Power Parity Indices**



**Figure 5: Average Real Wages of Permanent Residents by Aggregated Regions**

Total remittances received by the home country were obtained for 136 regions from the migrant transfers and workers compensation from the IMF balance of payments statistics. Ratha (2004) and Kapur and McHale (2005) point to numerous problems with misreporting and under-reporting of remittances, and therefore argue that both these definition should be included in the remittances data.

Remittances were then allocated across source regions to determine bilateral remittances by assuming a constant share of remittances to income. The resulting remittances paid per migrant are also shown in Figure 6. North America, Europe, Australia, Japan and parts of the Middle East have the highest remittances out per person, reflecting the higher wages earned by migrant labor there, and to a lesser extent the high numbers of people from high remitting regions.

**Figure 6: Remittances Paid by host Region per migrant**
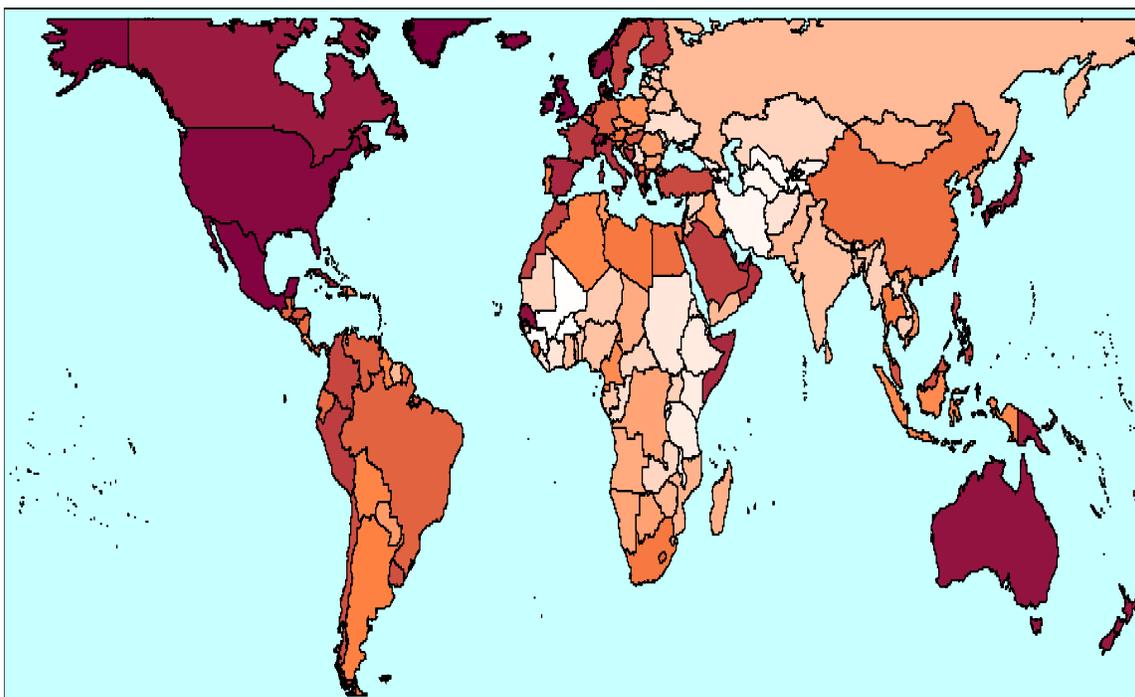


Note: The darker the color the larger the remittances paid per migrant.

The GTAP 8 Data Base does not include remittances; their inclusion in the database affects the income of both the home and host regions.[9] Moreover, since income earned must equal income spent, this change in income in the region must also affect expenditure in the region. In the GTAP Data Base savings are the residual, and hence these are adjusted in each region to ensure income equals expenditure. In countries where (net) remittances are received, income and saving will rise; in those where (net) remittances flow out of the country, income and savings fall. This is not to say that all remittance income is saved; but when the underlying GTAP Data Base was constructed and remittances were excluded from income, and savings were determined as the residual between income and consumption, hence artificially lowering savings in the standard GTAP Data Base. Adding remittances therefore improves the estimate of savings.

---

[9] We assume that all other income (from capital, land et cetera) accrues to permanent residents of the region and not to migrant labor.

# 5. Conclusion

The lack of data on international migration has been a severe impediment to the analysis of temporary and permanent migration between countries. Here we review some of the recent efforts to improve the data and use these data to construct a globally consistent database of bilateral population, labor by skill, wages, and remittances which can be used for analysis of migration issues. This database relies heavily on recent work undertaken by Özden, Parsons, Schiff and Walmsley (2011) to construct a global bilateral matrix of foreign born populations; and by Docquier et al. (2010) on the education levels of migrant labor have significantly improved the data available for analysis.

# *References*

Borjas, George (2000) "Immigration and the Food Stamp Program" JCPR Working Papers121, Northwestern University/University of Chicago Joint Center for Poverty Research.

Docquier, F., and A. Markouk (2004) "International Migration by Educational attainment (1999-2000) – Release 1.0" World Bank Policy Research Working Paper 3381.

Docquier, F., A. Markouk, C. Özden, and C. R. Parsons (2010) "Geographic, gender and skill structure of international migration" Report written for the Economic Research Forum.

Harrison, A., T. Britton, and A. Swanson (2003) "Working abroad – the benefits from nationals working in other economies" OECD.

IMF (2012), Balance of Payments Statistical Yearbook. Accessed online.

International Labor Organization (2012) LABORSTA Labour Statistics Database, Accessed online on 05/01/2012

Kapur, D., and J. McHale (2005) "Give us your best and brightest: A global hunt for talent and its impact on the developing world" Center for Global Development.

Liu, J., N. van Leeuwen, T. Thanh Vo, R. Tyers and T. W. Hertel, (1998) "Disaggregating Labor Payments by Skill Level in GTAP" GTAP technical paper No. 11, Center for Global Trade Analysis, Purdue University, West Lafayette, IN, USA.

Mattoo, A., I.C. Neagu, C. Özden, (2005) "Brain Waste? Educated Immigrants in the US labor market" World Bank Policy Research Working Paper, 3581.

Narayanan, G., A. Aguiar, and R. McDougall, Eds. (2012) Global Trade, Assistance, and Production: The GTAP 8 Data Base, Center for Global Trade Analysis, Purdue University.

OSSA-UN, 2005, Report on Resource Flows to Africa: An Update on Statistical Trends.

Özden, C., C. Parsons, M. Schiff, and T.L. Walmsley (2011) "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000" *The World Bank Economic Review*.

Parsons, C. R., R. Skeldon, L. A. Winters and T. L. Walmsley (2005) "Quantifying the international bilateral movements of migrants" Presented at the 8th Annual Conference on Global Economic Analysis, Lübeck, Germany, June 9-11.

Ratha, D. (2004) "Understanding the Importance of Remittances" Migration Policy Institute, http://www.migrationinformation.org/Feature/display.cfm?id=256

Walmsley, T. L.; S. Ahmed, and C.R. Parsons (2005) "The GMig2 Data Base: A Data Base of Bilateral Labor Migration, Wages and Remittances" GTAP Research Memorandum, 6; Center for Global Trade Analysis, Purdue University, Indiana, USA.