China and India's international migration within the Asia Pacific region: Insights from indirect estimation for the period 2000 to 2015

Jasmine Trang Ha, James Raymer, and Qing Guan

30 October 2019 (Draft)

Abstract

China and India are the two top origins and destinations for international migrants, yet the absence of migration flow data have hindered knowledge about how these two countries are connected to other origins and destinations within the Asia-Pacific region. This paper fills in the gap by estimating annual international migration flows for China and India between 2000 and 2015. Using a multiplicative component model framework, we borrow migration flow data from 30 other countries in the world and covariate information to produce out-of-sample predictions of total immigration and emigration. Bilateral flows are then estimated using a range of observed auxiliary information such as trade and remittance flows. The result is a complete and consistent account of migration flows amongst 53 countries that allows one to examine how migration has coincided with rapid demographic and economic change in China and India.

1. Introduction & Background

China and India are two most common places of birth for international migrants globally. According to the latest tally from the United Nations, over 40 million migrants in 2017 originated from China or India. Recent research has also documented the two countries' rising importance as destinations for returned nationals as well as foreign workers, particularly those in the broader Asia-Pacific region (Martin 2009). However, migration flow patterns from and to these two countries remain largely unknown because data are unavailable for cross-national comparison (Charles-Edwards et al., 2016; Hugo, 2005; Iredale et al., 2003).

Neither country has public data on total international migration flows and very few studies attempted to report concrete numbers. Luo (2003) referred to unpublished data he gathers from the Overseas Chinese Affairs Office and the Ministry of Foreign Affairs and stated that annual emigration from China is estimated to be 500,000. The date of the statistics, however, was not given. For India, both migration flow and border control statistics are largely absent. Studies of international migration in India tend to make use of immigrant population statistics derived from migrants' place of birth (Khadria 2001; Parida and Raman 2016).

In this paper, we seek to answer a single question: How many migrants move to/from China and India annually in the 2000-2015 period? With the focus on annual migration flow, our goal is to develop estimates of country-to-country migration flows which encompass all possible types of migration. Using the multiplicative component model framework (Raymer 2007; Willekens, 1982, 1983), we estimate annual bilateral migration flows amongst 53 Asia-Pacific countries from 2000 to 2015, borrowing data from 30 countries and auxiliary information including population sizes, demographic and economic conditions, as well as bilateral relationships such as trade and immigrant population. Our set of Asia-Pacific countries include countries and states in East Asia, South-East Asia, South Asia, Oceania and Pacific Islands, as well as Canada and the United States which have strong migration connections with the Asia-Pacific. The full list of Asia-Pacific countries can be found in Appendix A.

The major advantage of our estimation procedure is that it aligns with a multi-regional (versus uni-regional) approach to understand migration flows and migration patterns (Raymer, Willekens and Rogers 2018). With insight from spatial demography, our approach focuses on the dynamics and interactions of population change across space. In practice, estimating migration flows in a 54 by 54 two-way origin-destination table means that all flows are interdependent and connected through a system.

Our estimated annual migration flows thus represent an important first step in overcoming the problems of missing and incomplete data in China and India, as well as the broader Asia-Pacific region. As international migration represents a major source of demographic and social change for countries throughout the world, a concrete set of annual migration flow estimates would enhance our ability to study these changes and to understand the role of migration policies in determining the patterns.

2. Estimation methods

2.1 Migration data

In order to estimate migration flows effectively, some understanding of the different types of migration and migration data are needed. Migration data may be divided into two main types: immigrant populations and migration flows (Raymer, 2017). Immigrant population data are usually collected by censuses with questions on country of birth. They are abundantly available and represent the number of persons currently residing outside their country of birth at a particular point in time. These type of migration data provide useful information about the composition of migrants residing in particular countries but do not indicate when or how many migrants arrived during specific periods of time.

Migration flow data are collected from administrative data sources as the number of international arrivals or departures within a period of time (usually one year) or from survey or census questions on place of current residence by place of previous residence at some fixed point in time in the past (e.g., one-year ago). They are normally assumed to represent a change in the country of usual residence, following the United Nations (1998) recommendations. However, in practice these data are measured inconsistently and mostly unavailable. Moreover, there are two main types of migration flow data, consisting of flows by (previous/next) country of residence and by country of citizenship. Despite the importance of both measures, both happen to be inconsistent with the measurement of immigrant population stock data. In order to track changes in immigrant populations, one would need immigration and emigration flows by country of birth. For the purposes of our study, we are primarily interested in obtaining annual information on the migration flows by (previous/next) country of residence because of their usefulness in studying the levels and patterns of population redistribution across Asia-Pacific countries. They also provide a better understanding of past patterns and a more precise prediction of future trends (Abel forthcoming2018; Willekens et al., 2016).

2.2 Strategies to overcome missing and incomplete migration data

Overcoming issues with international migration data is a long-standing problem going back to the origins of coordinated international data collection (Bilsborrow et al., 1997; United Nations, 1949). Methods for harmonising across different definitions of migration and estimates of missing flows between countries, on the other hand, are relatively recent and have focused mainly on European countries, where data are more abundant (see, e.g., Abel, 2010; De Beer et al., 2010; Poulain, 1993; Raymer, 2008; Raymer et al., 2011, 2013). There have been attempts to estimate global flows of migration using gravity models (e.g., Cohen et al., 2008) and immigrant population stocks (e.g., Abel, 2013, forthcoming2018; Azose and Raftery, 2019). However, the plausibility of estimates resulting from these works are not clear because they are based on raw reports on migration that have not been harmonised or immigrant population stock data that underrepresent the annual flow aspects of migration, respectively.

The reasons why international migration data are so problematic and unreliable have to do with several factors. First, the process of migration itself is a rare event—the vast majority of people do not change their country of usual residence in any given year. So, any attempt to quantify those that do change their country of residence require specialised data collection systems to identify migrants from stayers. Second, most data sources are not designed to collect data for purposes of studying migration processes or demographic change. Instead, migration data are often a by-product of an administrative process used to regulate entry of non-citizens. This results in inconsistencies, coverage differences and underreporting. Third, countries utilise different concepts of what a migration represents and, although the United Nations (1998) provided some recommendations for measurement, hardly any countries abide by them.

To overcome the many obstacles concerning migration flow data, there are several approaches that can be utilised to provide synthetic estimates of the flows (Abel, 2010; De Beer et al., 2010; Raymer et al., 2011, 2013). These approaches tend to make use of the available (good quality) data and covariate information related to migration. For example, Raymer et al. (2013) used Bayesian inference to estimate migration flows in Europe by controlling for different measurement aspects and estimating missing data using a spatial interaction model that included covariates, such as population size in the origin and destination countries, immigrant population stocks and economic indicators. Moreover, they were able to utilise both sending and receiving country data for which some origin-destination flows could be compared and used as a basis for data harmonisation.

2.3 *The multiplicative component model framework*

We utilise the multiplicative component model framework proposed by Willekens (1982, 1983) for modelling internal migration and by Raymer (2007) for international migration. The model framework estimates migration flows, representing two-way (origin by destination) contingency tables (see Table 2), where the cells represent counts of persons, n_{ij}, who move from origin i to destination j in a given year. The methodology makes a distinction between cell counts (n_{ij}) and marginal totals, i.e., the total number of out-migrants from each region (n_{i+}), the total number of in-migrants to each region (n_{+j}) and the overall level of migration (n_{++}).

Specifically, the multiplicative component model decomposes a migration flow, n_{ij}, into four multiplicative components representing an overall level component, T, an origin main effect, O_i, a destination main effect, D_j, and an origin-destination interaction effect, OD_{ij}. The component T denotes the total number of migrants (i.e., n_{++}), O_i is the proportion of all migrants leaving from area i (i.e., n_{i+}/n_{++}) and D_j is the proportion of all migrants moving to area j (i.e., n_{+j}/n_{++}). The interaction component OD_{ij} is defined as $n_{ij}/[(T)(O_i)(D_j)]$ or the ratio of observed migration to expected migration (for the case of no interaction).

The idea behind the multiplicative component method is that tables of migration flows can be decomposed into various hierarchical structures, not all of which are necessary for understanding or for producing accurate predictions. If certain (important) structures are unavailable, they can be imputed or 'borrowed' from auxiliary data sources. The advantage of this model is that the missing margin totals (n_{i+} and n_{+j}) can be estimated based largely on the known correlation with the populations 'at risk' of sending and receiving international migration. For instance, we expect larger flows of migration both from and to larger populations. These total flows may then be distributed across various destinations using auxiliary information about the connectivity between places (as a proxy for OD_{ij}), such as distances between countries, the presence of immigrant populations, and bilateral trade flows. This model framework, similar to the generation and distribution model proposed by Willekens and Baydar (1986), has been shown effective for overcoming problems of severe missing internal and international migration flows data (De Beer et al., 2010; Raymer et al., 2011). Further, as the multiplicative component model requires estimating three components (n_{i+}, n_{+j} and OD_{ij}) separately, we can further assess the plausibility of each component independent of the other components.

Our strategy to model the international migration flows amongst Asia-Pacific countries from 2000 to 2015 includes three steps, as outlined in Figure 1 below. First, we use two regression models with input data from non-ASEAN countries to generate predictions for total immigration and total emigration each Asia-Pacific country. Predictions for the Rest-of-world category is calculated using immigrant population data from the United Nations. This step is used to generate estimates for the overall level effect T, origin main effect O_i and destination main effect D_j. To enhance the accuracy and plausibility of our estimates, we further use reported data in Australia, New Zealand, and South Korea to benchmark the size of total inflows and outflows. In the second step, we use a regression model to estimate the OD_{ij} interaction components. Finally, we estimate n_{ij} by rescaling the values to match the estimated margins produced in Step 1 above through iterative proportional fitting (IPF). More details on our estimation procedure can be found in Appendix B.

--- FIGURE 1 ABOUT HERE---

Our estimation procedure is based on two types of data: migration data and covariate data. Migration theories suggest relationships between migration and a country's characteristics, such as population size or GDP per capita. Thus, we build regression models from countries where both migration data and covariate data are available. Then, using the regression coefficients, we generate migration estimates for Asia-Pacific countries where only covariate information is available. Table 1 indicates our data sources. High-quality migration data are only available in selected European countries, where much work has been carried out in comparing and assessing data quality (see, e.g., De Beer et al., 2010; Kupiszewska and Nowok, 2008; Poulain et al., 2006; Raymer et al., 2011, 2013). While there are obvious differences between Europe and Asia-Pacific, prior studies have suggested that European data could be used as a "bronze standard," which have sufficient quality for capturing migration patterns and the relationships between migration and other social-economic processes (Azose and Raftery 2019).

--- TABLE 1 ABOUT HERE---

3. Results

Following our 3-step estimation procedure, the final estimates indicate rapid rises in both emigration and immigration in China and India in the 2000-2015 period. Figure 2 outlines the changes in emigration and immigration in the two countries over the 16-year estimation period.

--- FIGURE 2 ABOUT HERE---

In China, we estimate that there are under 2 million immigrants and emigrants in 2000, and these number rise sharply to over 4.5 million emigrants and over 9.3 million immigrants in 2015. A similar pattern of increase in found in India. However, while immigration is estimated to become more prominent in China, our estimates suggest that India is still largely an emigration country, as the size of emigration flows are two to three times larger than the size of immigration flows.

Further, the estimated emigration and immigration rate as well as the ratio of emigration over immigration suggest that our estimates are plausible in a demographic sense. In both countries, immigration and emigration account for less than 3% of annual population change. As an additional step to consider the plausibility of our estimates, we assess the marginal totals using the demographic balancing equation and information about the country's natural increase obtained from the World Development Indicators database. We use our estimated emigration and immigration values in combination with reported natural increase rates to age each country's population forward. We then compare the difference between our projected population sizes and the reported ones. As shown in Panel A of Figure 3, the difference falls between ±2% of reported population sizes for 90% of the estimated country-year. The maximum difference is -5% and the maximum difference is +10%. The outliers are few, and the tends to be countries with very small populations: Federated States of Micronesia (FSM), Macau (MAC), Tonga (TON), and Western Samoa (WSM).

--- FIGURE 3 ABOUT HERE---

We further display the difference for selected countries in Panel B of Figure 3, including Australia (AUS), Canada (CAN), China (CHN), India (IND), South Korea (KOR), and New Zealand (NZL). Compared to countries where population data and vital statistics on birth and death are more accurate, our migration estimates for China and India are doing quite well in matching the changes in birth and death and coming up to a very close account of population change.

Our full bilateral migration flow estimates can be examined using circular flow plots similar to Figure 4, where we display two plots, one for the year 2000 and the other for the year 2015. As indicated by the relative thickness of the flows, we can see that there are much more migrants in the systems in year 2000 compared to year 2015. One can also examine a specific flow, for example from India (IND) to China (CHN) and note that the flow has become larger in 2015.

--- FIGURE 4 ABOUT HERE---

The estimates also help understand regional patterns, for example, of migration between China and India and countries in South-East Asia. Existing studies suggested that inflows from China to South-East Asia were considerably large in the 1990-2010 period, particularly towards countries with abundant economic opportunities such as Thailand, Singapore, and Malaysia (Wong 2013; Xiang 2012). South-East Asian countries are also important “stepping-stones” for Indian and Chinese migrants who subsequently migrated to more economically advanced societies, such as the United States or Canada (Min and Park 2014). As shown in Figure 5, our estimates partially confirm these suggestions, as the number of migrants from China and India moving into South-East Asia both increase over time in the 2000-2015 period. However, in China, there is a slow-down in outflows towards South-East Asia since 2004. Our estimates also draw attention to the counterflows, showing large inflows from South-East Asia as well.

--- FIGURE 5 ABOUT HERE---

4. Discussions

In this paper, we have designed a strategy to estimate the migration flows in the absence of any migration data amongst 53 Asia-Pacific countries. This strategy used a hierarchical and multiplicative component model to derive the estimates. The assumptions were based on the beliefs that total immigration and emigration flows could be estimated using the associations between covariate information and migration flows measured in other countries elsewhere in the world (mostly Europe). These associations were then applied to covariate information available for ASEAN countries.

Once annual estimates of the total immigration and emigration flows were obtained for each Asia-Pacific country from 2000 to 2015, we then allocated the origin-destination flows based on our estimates of the ratios of observed to expected flows. The final set of numbers were then assessed according to their demographic plausibility and what might be expected based on migration theory. Our estimates show that large numbers of people are migrating amongst Asia-Pacific countries each year. Our estimated 2015 flow table resulted over 52 million persons changing their country of residence. For China and India, our estimates suggest very sharp increases in both emigration and immigration in the two countries over the 16-year period. The estimates also help examining regional migration patterns, for instance, between China and South-East Asia.

Further, our estimation is based on previous efforts in Europe and takes into account covariate information known to be related to migration. Thus, our estimates are aligned with our expectations that population size is a major driver of international migration and that variables, such as shared borders, GDP per capita, percentage urban, female life expectancy and trade are all influencing factors.

By estimating the migration flows for the Asia-Pacific region, we learned that one can overcome severe data limitations to provide much needed information on how people are likely to be migrating amongst these ten countries over time. We also learned that more sophisticated validations strategies are needed to assess the quality of the migration flows estimates. A great follow-up strategy would be to solicit experts' views on the levels and patterns of the estimated flows, similar to the strategy carried out in the IMEM project (Wiśniowski et al., 2013).

There are two main contributions provided in this paper. The first is an estimated set of annual migration flows covering 16 years for all 53 Asia-Pacific nations. The estimated flows are aligned with the United Nations (1998) recommendations on the measurement of migration flows. Prior to this research, no such information was available and most of our understanding was based on qualitative or immigrant population stock data. The second contribution is the estimation and validation framework for estimating migration flows in the context of extremely limited information. While both aspects could certainly be improved, it sets the foundation for further research on overcoming data limitations for studying international migration. We hope our efforts will inspire others to provide better understandings of demographic processes occurring in areas of great demographic and economic change.

References

Abel GJ (2010) Estimation of international migration flow tables in Europe. *Journal of the Royal Statistical Society, Series A* 173(4): 797–825.

Abel GJ (2013) Estimating global migration flow tables using place of birth data. *Demographic Research* 28: 505–546.

Abel GJ (2018) Estimates of global bilateral migration flows by gender between 1960 and 2015. *International Migration Review,* 52(3): 809-852.

Asis MMB and Battistella G (2018) Irregular migration in Asia: Are new solutions in sight? In: Liu-Farrer G and Yeoh BSA (eds) *Routledge handbook of Asian migrations*. London: Routledge, pp. 277-287.

Azose JJ and Raftery AE (2019) Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences* 116(1): 116-122.

Baas M (2018) Temporary labour migration. In: Liu-Farrer G and Yeoh BSA (eds) Routledge *Handbook of Asian migrations*. London: Routledge, pp. 51-63.

Bilsborrow RE, Hugo G, Oberai AS and Zlotnik H (1997) *International migration statistics: Guidelines for improving data collection systems.* Geneva: International Labour Office.

Charles-Edwards E, Muhidin S, Bell M, and Zhu Y (2016) Migration in Asia. In: White MJ (ed) *International handbook of migration and population distribution.* Dordrecht: Springer, pp. 269-284.

Cohen JE, Roig M, Reuman DC and GoGwilt C (2008). International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences* 105: 15269–15274.

De Beer J, Raymer J, Van der Erf R and Van Wissen L (2010). Overcoming the problems of inconsistent international migration data: A new method applied to flows in Europe. *European Journal of Population* 26: 459-481.

Hugo G (2005) The new international migration in Asia: Challenges for population research. *Asian Population Studies* 1(1): 93-120.

Huguet JW (2016) APMJ – Chronicling 25 years of migration in Asia and the Pacific. *Asian and Pacific Migration Journal* 25(4): 357-378.

Iredale R, Guo F and Rozario S (2003). Introduction. In: Iredale R, Guo F and Rozario S (eds) *Return migration in the Asia Pacific*. Cheltenham: Edward Elgar, pp. 1-25.

Khadria, B (2001). Shifting paradigms of globalization: the twenty-first century transition towards generics in skilled migration from India. *International Migration*, 39(5):45-71.

Kupiszewska D and Nowok B (2008) Comparability of statistics on international migration flows in the European Union. In: Raymer J and Willekens F (eds) *International migration in Europe: Data, models and estimates.* Chichester: Wiley, pp. 41-71.

Luo K (2003). China's changing international migration trends. In Robin R. Iredale, Charles, Hawksley and Stephen Castles (Eds) *Migration in the Asia Pacific: Population, Settlement and Citizenship Issues.* Cheltenham: Edward Elgar Publishing Limited. pp 293-.

Min PG and Park SS (2014). Twice-Migrant Chinese and Indians in the United States: Their Origins and Attachment to Their Original Homeland. *Development and Society,* 43(2):381–401.

Özden C, Parsons C, Schiff M and Walmsley TL (2011) Where on earth is everybody? The evolution of Global Bilateral Migration, 1960-2000. *World Bank Economic Review,* 25(1):12-56. Data retrieved from https://datacatalog.worldbank.org/dataset/global-bilateral-migration-database

Parida, JK and Raman, KR (2016). India: Rising trends of international and internal migration. In Anna Triandafyllidou (Ed.) *Handbook of Migration and Globalisation.* Cheltnham: Edward Elgar Publishing Limited. PP. 226-246.

Park J-H, Dumont J-C and Baruah N (Eds.) (2015) *Building human capital through labour migration in Asia.* Tokyo: Asian Development Bank Institute.

Poulain M (1993) Confrontation des Statistiques de migrations intra-europe´ennes: Vers plus d'harmonisation? *European Journal of Population* 9(4): 353–381.

Poulain M, Perrin N and Singleton A (Eds.) (2006) *THESIM: Towards harmonised European statistics on international migration.* Louvain-la-Neuve: UCL Presses.

Raymer J (2007) The estimation of international migration flows: A general technique focused on the origin-destination association structure. *Environment and Planning A* 39: 985-995.

Raymer J (2008) Obtaining an overall picture of population movement in the European Union. In: Raymer J and Willekens F (eds) *International migration in Europe: Data, models and estimates.* Chichester: Wiley, pp. 209–234.

Raymer J (2017) Measuring flows of international migration. *IZA World of Labour* (354): 1-10.

Raymer J, De Beer J and Van der Erf R (2011) Putting the pieces of the puzzle together: Age and sex-specific estimates of migration amongst countries in the EU/EFTA, 2002-2007. *European Journal of Population* 27: 185-215.

Raymer J, Willekens F and Rogers A. 2018. Spatial demography: A unifying core and agenda for further research. *Population, Space and Place* 25(4): e2179.

Raymer J, Wiśniowski A, Forster JJ, Smith PWF and Bijak J (2013) Integrated modelling of European migration. *Journal of the American Statistical Association* 108(503): 801-819.

Shen W (2010). China in the global migration order—historical perspectives and new trends. *Asia Europe Journal* 8(1): 25-43.

Skeldon R (2006) Recent trends in migration in East and Southeast Asia. *Asian and Pacific Migration Journal* 15(2): 227-293.

Tomas PA and Summers LH (2009) *Migrants count: Five steps toward better migration data*. Report of the Commission on International Migration Data for Development Research and Policy, Center for Global Development, Washington DC.

United Nations (1949) Problems of migration statistics. *Population Studies* No. 5, Department of Social Affairs, United Nations, New York.

United Nations (1998) Recommendations on statistics of international migration. *Statistical Papers Series* M, No. 58, Rev.1, Department of Economic and Social Affairs, Statistics Division, United Nations, New York.

United Nations (2013) *International Migration Report 2013*. United Nations, New York.

Willekens F (1982) Multidimensional population analysis with incomplete data. In: Land K and Rogers A (eds) *Multidimensional mathematical demography.* New York: Academic Press, pp. 43-111.

Willekens F (1983) Log-linear modelling of spatial interaction. *Papers of the Regional Science Association* 52: 187-205.

Willekens F (1994) Monitoring international migration flows in Europe: Towards a statistical data base combining data from different sources. *European Journal of Population* 10: 1–42.

Willekens F (2019) Evidence-based monitoring of international migration flows in Europe. *Journal of Official Statistics* 35(1): 231-277.

Willekens F and Baydar N (1986) Forecasting place-to-place migration with generalized linear models. In: Woods R and Rees P (eds) *Population structures and models: Developments in spatial demography*. London: Allen & Unwin, pp. 203–244.

Willekens F, Massey D, Raymer J and Beauchemin C (2016) International migration under the microscope. *Science* 352(6288): 897-899.

Wiśniowski A, Bijak J, Christiansen S, Forster JJ, Keilman N, Raymer J and Smith PWF (2013) Utilising expert opinion to improve the measurement of international migration in Europe. *Journal of Official Statistics* 29(4): 583-607.

Wong D (2013) Introduction: The New Chinese Migration to Southeast Asia. *Asian and Pacific Migration Journal* 22(1):1–6.

Xiang B (2012) Predatory Princes and Princely Peddlers: The State and International Labour Migration Intermediaries in China. *Pacific Affairs*, 85(1):47-68.

Table 1. Sources of data

Type of data	Source of data	Year
Migration data	IMEM data base for 30 European countries (Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, the Netherlands and the United Kingdom)	2002-2008
Covariates predicting total immigration and emigration		1999-2014
Population size	World Development Indicator	
GDP adjusted for PPP (Purchasing power parity)	World Development Indicator	
% Foreign-born	United Nations Immigrant Population Database (2017 revision)	
Old-age dependency ratio	World Development Indicator	
Unemployment rate	World Development Indicator	
% Urban population	World Development Indicator	
Female life expectancy	World Development Indicator	
Covariates predicting the interaction component (ODij)		2000-2015
Contiguity	Centre d'Etudes Prospectives et d'Informations Internationales (CEPII)	
Common language	CEPII	
Common colony	CEPII	
Immigrant Population data	United Nations Database (2017 revision)	
Bilateral Trade	UN Comtrade Database	

Figure 1. Estimation procedure using the multiplicative component model framework

Step 1. Estimate the margin totals (n_{i+} and n_{+j})

	A	B	C	Total
A				n_{i+}
B				n_{i+}
C				n_{i+}
Total	n_{+j}	n_{+j}	n_{+j}	n_{++}

\+

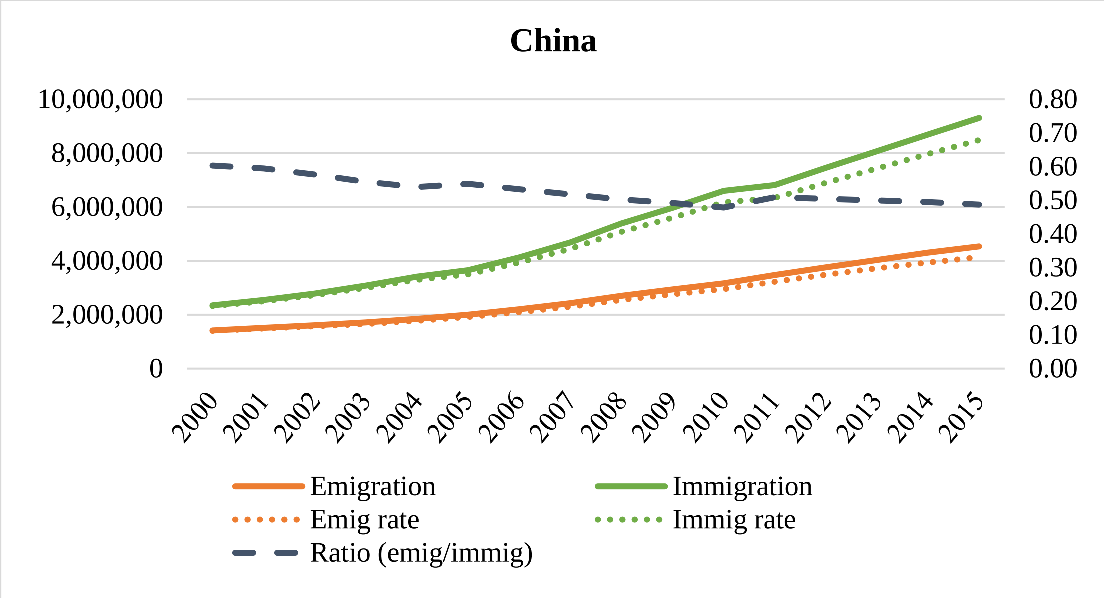
Step 2. Estimate the interaction component (OD_{ij})

	A	B	C
A	0	OD_{ij}	OD_{ij}
B	OD_{ij}	0	OD_{ij}
C	OD_{ij}	OD_{ij}	0

Step 3. Estimate n_{ij} using Iterative Proportional Fiting (IPF)

	A	B	C	Total
A	0	n_{ij}	n_{ij}	n_{i+}
B	n_{ij}	0	n_{ij}	n_{i+}
C	n_{ij}	n_{ij}	0	n_{i+}
Total	n_{+j}	n_{+j}	n_{+j}	n_{++}

Figure 2. Estimated total emigration and immigration in India and China, 2000-2015



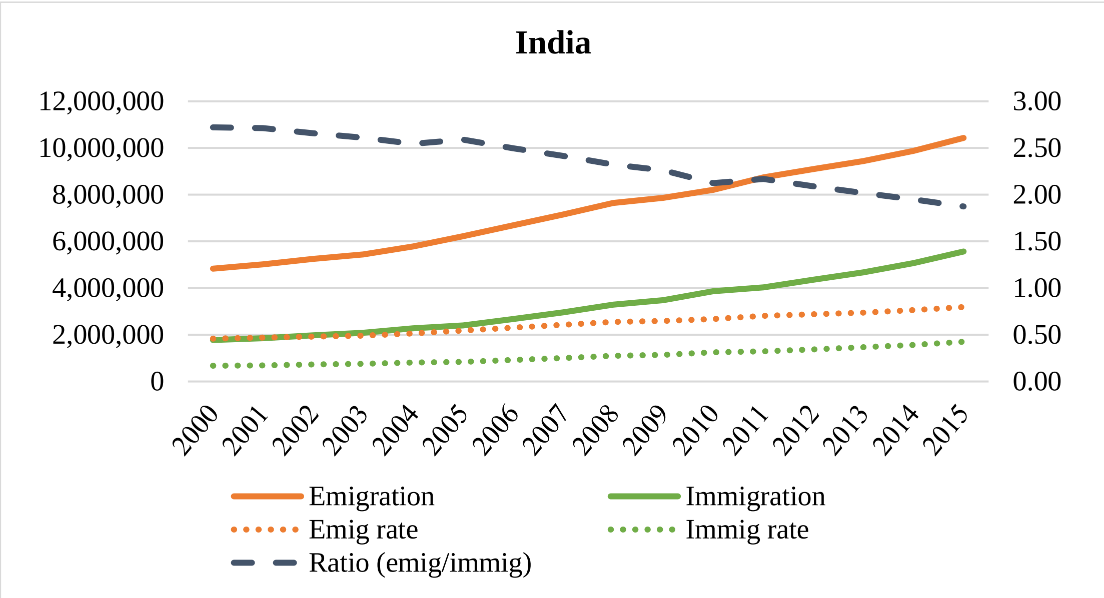
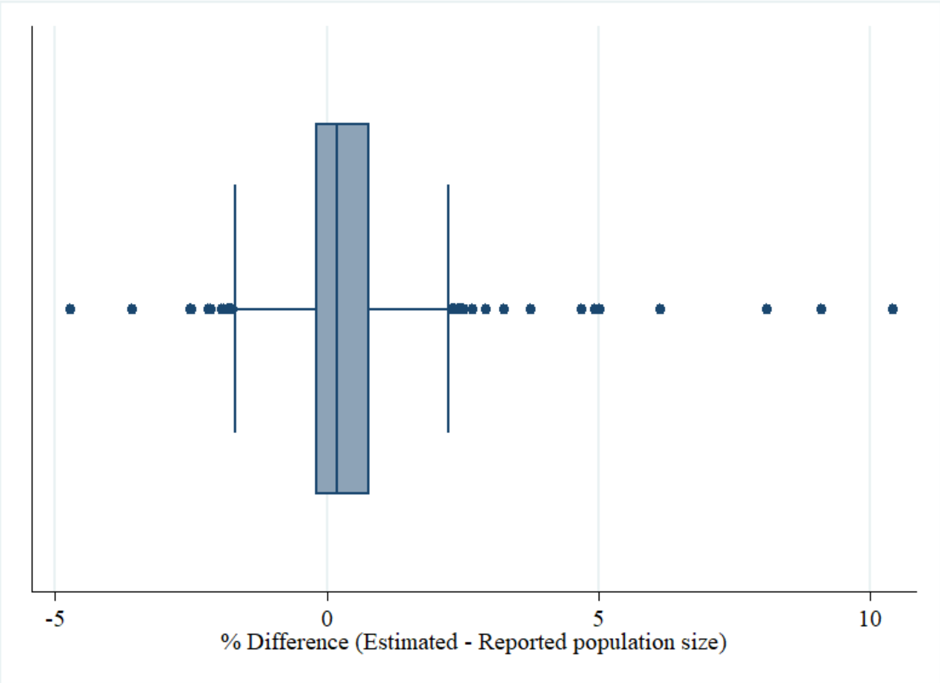


Figure 3. Plausibility checks using demographic balancing equation

Panel A. Overall difference (all country-years)



Panel B. Selected countries

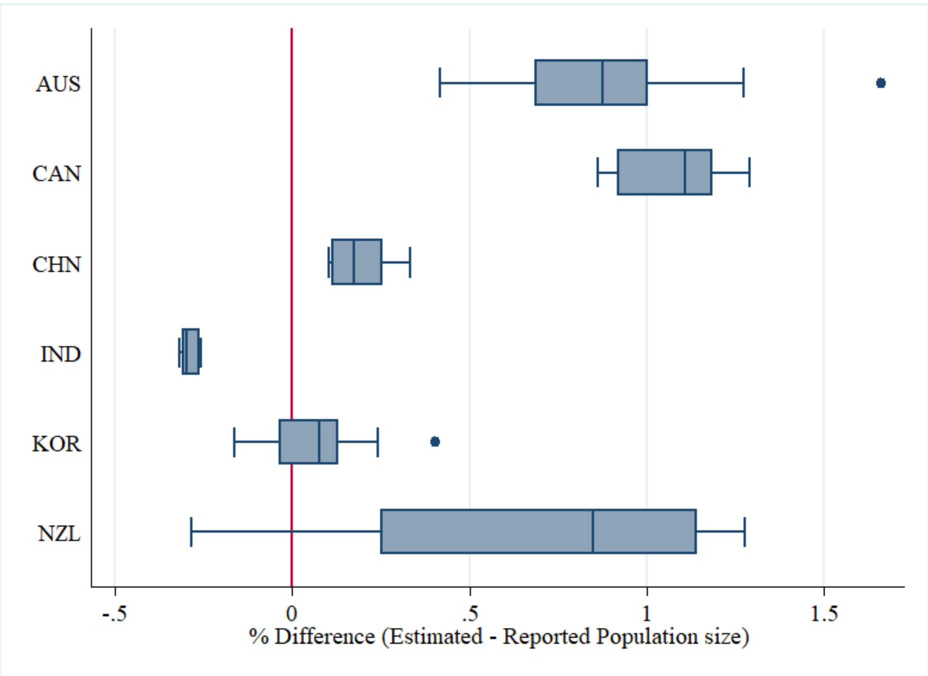
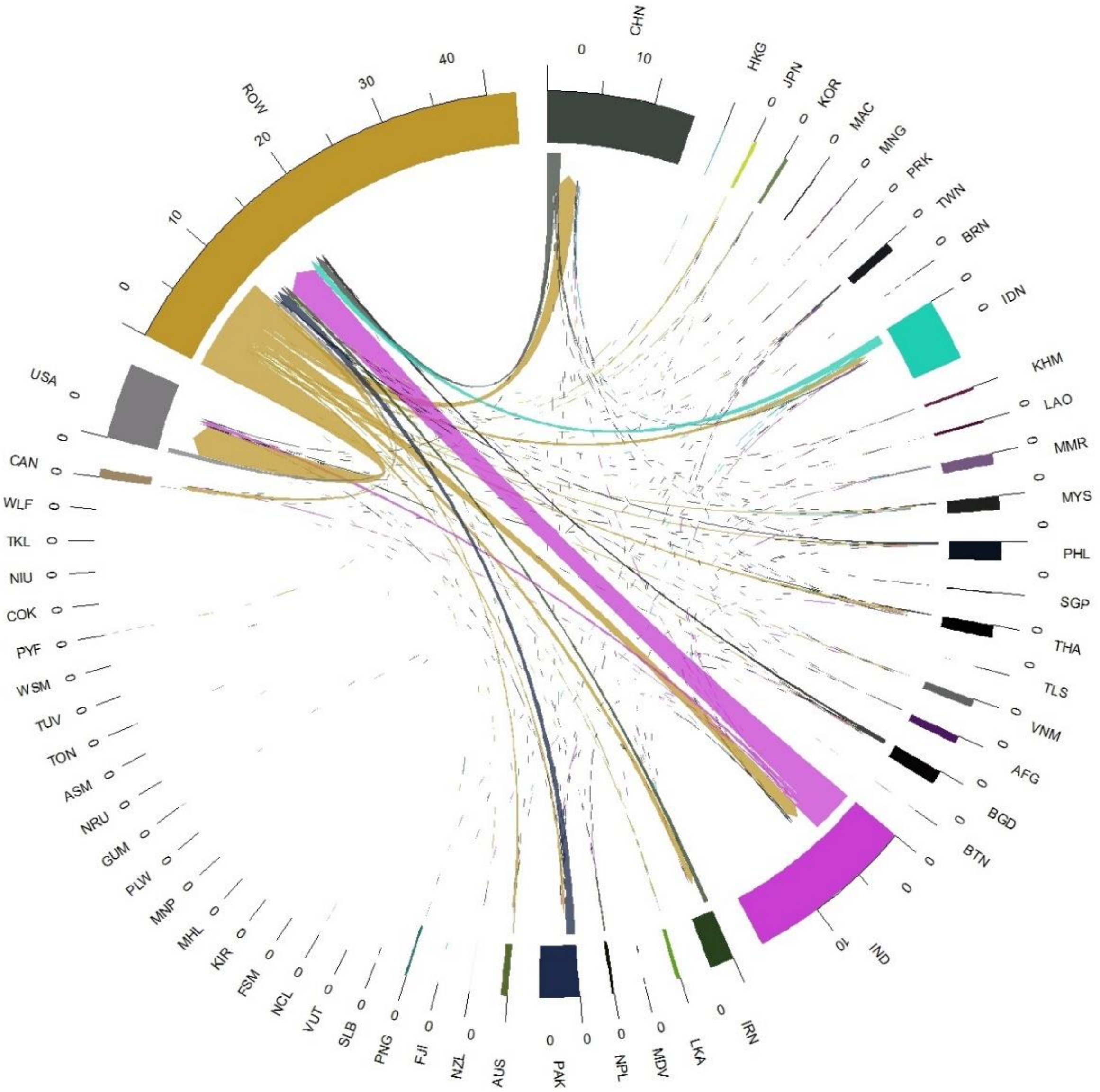
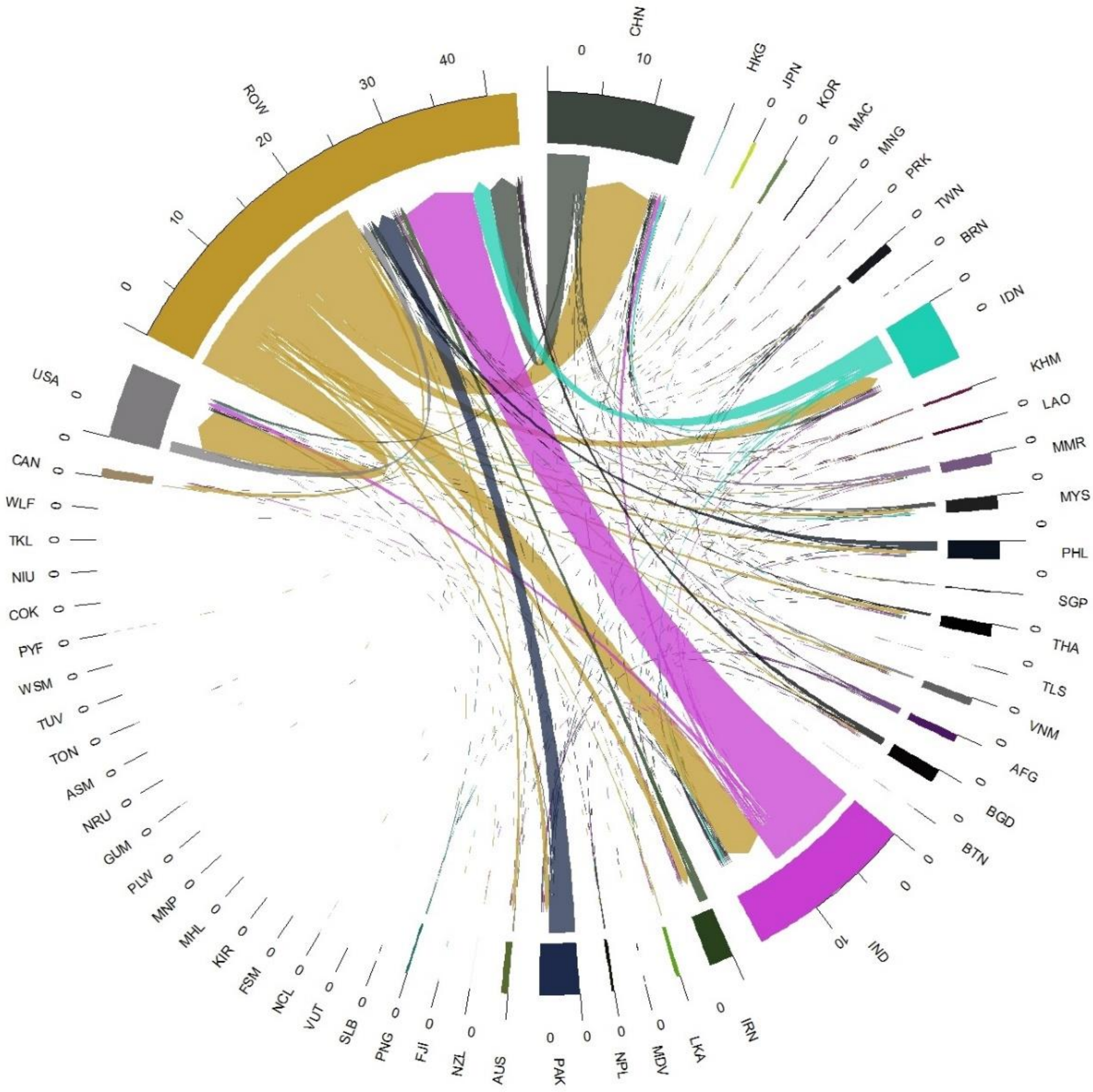


Figure 4. Bilateral flow estimates in 54 Asia-Pacific countries in 2000 and 2015.

Year 2000

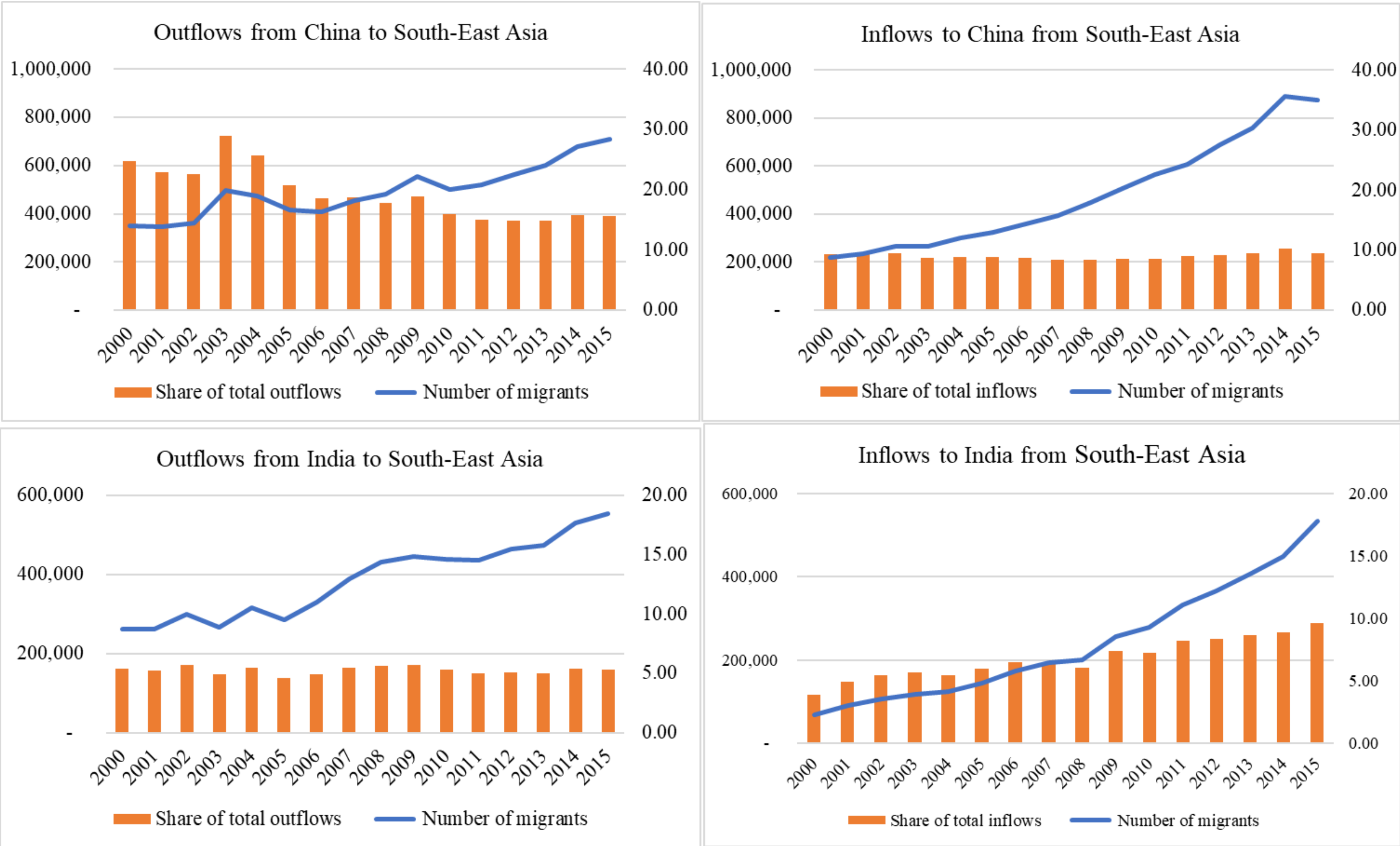


Year 2015



CHN
HKG
JPN
KOR
MAC
MNG
PRK
TWN
BRN
IDN
KHM
LAO
MMR
MYS
PHL
SGP
THA
TLS
VNM
AFG
BGD
BTN
IND
IRN
LKA
MDV
NPL
PAK
AUS
NZL
FJI
PNG
SLB
VUT
NCL
FSM
KIR
MHL
MNP
PLW
GUM
NRU
ASM
TON
TUV
WSM
PYF
COK
NIU
TKL
WLF
CAN
USA
ROW

Figure 5. Migration patterns between India, China, and South-East Asian countries



Appendix A. List of Asia-Pacific countries

Number	ISO3	Country name	World region
1	TWN	Taiwan	Eastern Asia
2	TKL	Tokelau	Polynesia
3	WLF	Wallis and Futuna Islands	Polynesia
4	COK	Cook Islands	Polynesia
5	NIU	Niue	Polynesia
6	NCL	New Caledonia	Melanesia
7	GUM	Guam	Micronesia
8	PYF	French Polynesia	Polynesia
9	NRU	Nauru	Micronesia
10	TUV	Tuvalu	Polynesia
11	MHL	Marshall Islands	Micronesia
12	PLW	Palau	Micronesia
13	MNP	Northern Mariana Islands	Micronesia
14	ASM	American Samoa	Polynesia
15	KIR	Kiribati	Micronesia
16	FSM	Micronesia (Federated States of)	Micronesia
17	PRK	Democratic People's Republic of Korea	Eastern Asia
18	CHN	China	Eastern Asia
19	HKG	China, Hong Kong Special Administrative Region	Eastern Asia
20	MAC	China, Macao Special Administrative Region	Eastern Asia
21	JPN	Japan	Eastern Asia
22	MNG	Mongolia	Eastern Asia
23	KOR	Republic of Korea	Eastern Asia
24	BRN	Brunei Darussalam	South-Eastern Asia
25	KHM	Cambodia	South-Eastern Asia
26	IDN	Indonesia	South-Eastern Asia
27	LAO	Lao People's Democratic Republic	South-Eastern Asia

28	MYS	Malaysia	South-Eastern Asia
29	MMR	Myanmar	South-Eastern Asia
30	PHL	Philippines	South-Eastern Asia
31	SGP	Singapore	South-Eastern Asia
32	THA	Thailand	South-Eastern Asia
33	TLS	Timor-Leste	South-Eastern Asia
34	VNM	Viet Nam	South-Eastern Asia
35	AFG	Afghanistan	Southern Asia
36	BGD	Bangladesh	Southern Asia
37	BTN	Bhutan	Southern Asia
38	IND	India	Southern Asia
39	IRN	Iran (Islamic Republic of)	Southern Asia
40	MDV	Maldives	Southern Asia
41	NPL	Nepal	Southern Asia
42	PAK	Pakistan	Southern Asia
43	LKA	Sri Lanka	Southern Asia
44	NZL	New Zealand	Australia and New Zealand
45	FJI	Fiji	Melanesia
46	PNG	Papua New Guinea	Melanesia
47	SLB	Solomon Islands	Melanesia
48	VUT	Vanuatu	Melanesia
49	WSM	Samoa	Polynesia
50	TON	Tonga	Polynesia
51	CAN	Canada	North America
52	USA	United States of America	North America
53	AUS	Australia	Australia and New Zealand

Appendix B. Details about the estimation procedure

In the first step, we use a regression models using migration and covariate data from 30 European countries to establish the relationships between different covariates and migration flows. Table A1 displays results from two sets of models, one predicting total emigration and the other predicting total immigration. When some covariates are missing for a country (e.g., Taiwan), we use an alternative model requiring less covariates. As indicated by the adjusted R-square, models E2, E3, I2, and I3 have less explanatory power compared to E1 and I1, yet they still account for over 90% of the variation in migration flows.

Table A1. OLS regression models predicting total immigration and total emigration

	Emigration			Immigration		
	Model E1	Model E2	Model E3	Model I1	Model I2	Model I3
Population size (lag, log)	0.869***	0.763***	0.733***	0.900***	0.928***	0.942***
	(0.026)	(0.031)	(0.031)	(0.028)	(0.026)	(0.026)
Dummy for small country^	-0.533***	-0.333***	-0.370***	-0.179*	-0.120	-0.104
	(0.064)	(0.090)	(0.091)	(0.076)	(0.068)	(0.070)
Adjusted GDP per capita (lag, log)	0.522***	0.049	0.160**	0.813***	0.941***	0.858***
	(0.110)	(0.069)	(0.060)	(0.120)	(0.051)	(0.045)
Year	0.062***	0.050***	0.047**	-0.026*	-0.026*	-0.022*
	(0.009)	(0.014)	(0.015)	(0.011)	(0.011)	(0.011)
% Foreign-born (lag)	0.026***	0.016**		-0.011*	-0.013**	
	(0.004)	(0.005)		(0.004)	(0.004)	
Old-age dependency ratio (lag)	-0.088***			-0.015		
	(0.007)			(0.010)		
Unemployment rate (lag)	-0.002			0.000		
	(0.007)			(0.007)		
% Urban population (lag)	0.004*			-0.011***		
	(0.002)			(0.002)		
Female life expectancy (lag)	-0.130***			0.061**		
	(0.015)			(0.018)		
Emigration rate				0.483***	0.488***	0.452***
				(0.073)	(0.052)	(0.052)
Constant	-119.554***	-101.701***	-96.248**	39.872	42.316	34.440
	(18.937)	(28.492)	(29.021)	(21.335)	(21.592)	(21.972)
Observations	210	210	210	210	210	210
R-squared	0.96	0.91	0.90	0.97	0.96	0.96
Adjusted R-squared	0.96	0.91	0.90	0.97	0.96	0.95

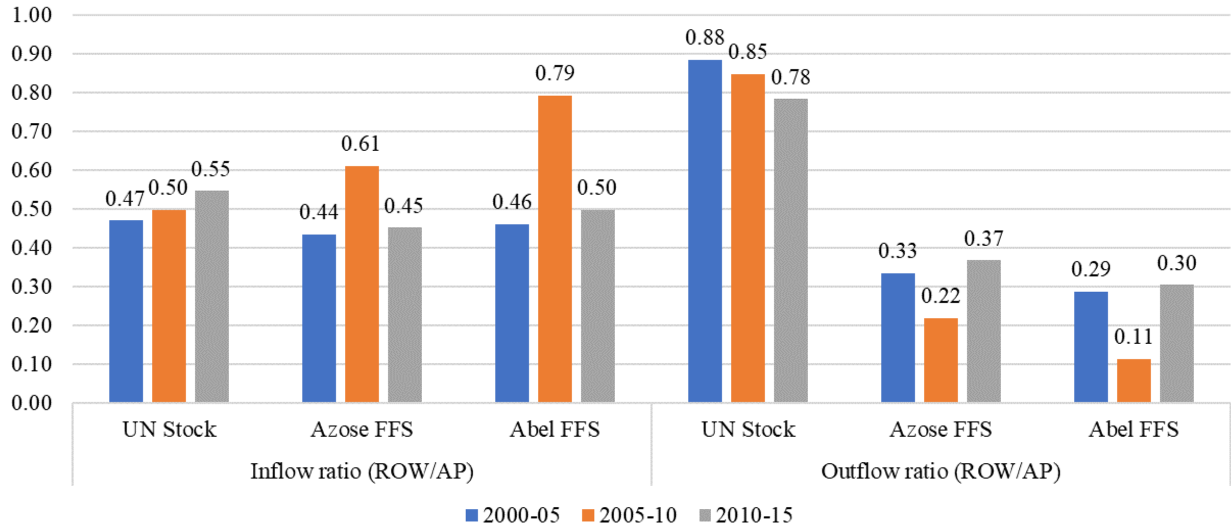
Notes: Standard errors in parentheses, * p<0.05; ** p<0.01; *** p<0.001, ^ indicates country with population size smaller than 6 million.

Consistent with prior research (Cohen et al., 2008; Jennissen, 2004), the results indicate that population sizes are strong predictors of annual emigration and immigration.

The variable on adjusted GDP per capita has a positive and statistically significant effect for predicting immigration and emigration. The coefficients for the percentage elderly exhibited negative and statistically significant effects for predicting emigration but not for immigration. The variable on percentage urban resulted in higher predictions for emigration flows but lower for immigration. Higher female life expectancy resulted in lower predicted values for emigration and higher predicted values for immigration. The year variable resulted in a negative and statistically significant effect for immigration but positive for emigration. Finally, the results from Model I1-3 indicate that countries with higher emigration rates would also have higher immigration flows. The adjusted R-squared values are higher than 0.90 for all models.

Having obtained the estimates of total immigration and emigration for each of the 53 Asia-Pacific (AP) countries, we then complete the two-way origin-destination table by calculating inflows and outflows to a residual Rest-of-world (ROW) category. We rely on other data to estimate the ratio of ROW/AP for both inflows and outflows. In Figure A1 below, we show how the ratios look like using three different sources of data: the UN immigrant population data (or UN Stock), flow-from-stock estimates by Abel (2018) and flow-from-stock estimates by Azose and Raftery (2019). In both flow-from-stock estimates, the ratio does not have a linear pattern, rather, the period 2005-10 tends to be quite different from the other two periods. We think that the sharp changes in ratio are not plausible, and therefore we elect to use the ratio derived from UN Stock data. For example, for the years 2000, 2001, 2002, and 2004, we multiply total inflows to Asia-Pacific countries by 0.47 to arrive at an estimate of inflows to ROW.

Figure A1. Ratios used for calculating Rest-of-world inflows and outflows



As our estimates of total immigration and total emigration are derived from two separate models, the levels of immigration and emigration might not be reliable. The difference becomes clear when we force them to fit the same grand total (T) of the two-way origin-destination table. In order to bring our estimates to a plausible level, we use reported migration data from countries with relatively good data quality to benchmark our estimate. The three countries are Australia (AUS), New Zealand (NZL), and South Korea (KOR).

Figure A2. Ratios of reported over estimated inflows and outflows for selected countries

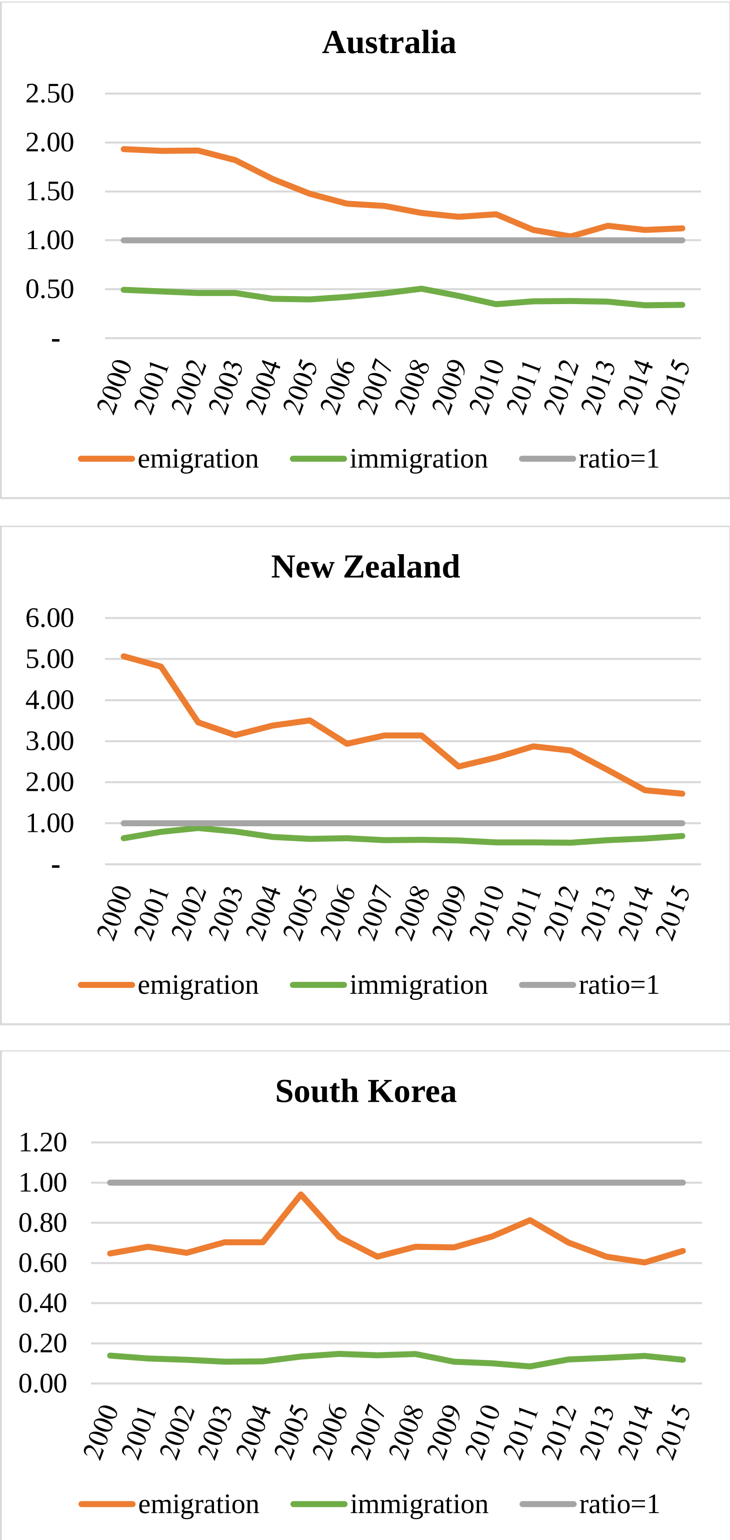


Figure A2 display the ratio of reported over our initial estimated inflows and outflows. Our estimates for emigration are lower than reported in both Australia and New Zealand, but they are slightly higher than reported data in South Korea. In terms of immigration, our estimates are always higher than the reported figure, and in South Korea, they are a lot higher than reported. The high over-estimation ratio for South Korea might be due to the fact that South

Korea are much more restrictive in its migration control policy, compared to European countries and to Australia and New Zealand. While South Korea has a steady inflow of labour migrants, their strict control policy means that labour migrants tend to leave after finishing their contracts rather than settling. Thus, despite its high attraction, South Korea does not gain as much immigration as Europe or Australia and New Zealand. Other countries in Asia-Pacific are found to have similar restrictive control policies, including Singapore, Taiwan, Hongkong, Japan, and Brunei (Asis and Battistella 2018; Baas 2018).

Our benchmarking procedure thus includes two steps. First, we benchmark all emigration totals to the average ratio of reported over estimated in Australia, New Zealand, and South Korea for all years. The average ratio is 1.73, thus we multiply the estimated total emigration by 1.73 to increase the levels of emigration. Second, we benchmark the countries with restrictive immigration control similar to South Korea to the average ratio of reported over estimated in South Korea. In effect, multiplying total immigration estimates to a ratio of 0.12 makes the estimates a lot smaller for South Korea, Singapore, Taiwan, Hongkong, Japan, and Brunei. Finally, for all other countries, we benchmark total immigration to the average ratio of reported over estimated in Australia and New Zealand, to a ratio of 0.53.

In the second step, we estimate the interaction component (or OD_{ij}) using a set of OLS regression models. Similar to the first step, here we also borrow data from 30 European countries in the 2002-2008 period, making a total of 30x29x8=6,960 observations. Then we use the coefficients to generate predicted values for the interaction component amongst pairs of Asia-Pacific countries. The model results are shown in Table A2. When certain covariates are missing, we use alternative model that contains fewer variables (Model 2 and Model 3).

Table A2. OLS regression models predicting the interaction terms

	Model 1	Model 2	Model 3
Bilateral migrant stock interaction	0.529***	0.522***	
	(0.009)	(0.008)	
Bilateral trade interaction	0.463***	0.402***	0.499***
	(0.012)	(0.011)	(0.009)
Migrant stock x Trade	-0.010***	-0.012***	
	(0.000)	(0.000)	
Year	0.039**	0.038*	0.020
	(0.015)	(0.015)	(0.020)
Contiguity	-0.783***		
	(0.121)		
Common official language	0.632***		
	(0.159)		
Common colonial history	-7.432***		
	(0.442)		
Previous colony-colonizer pair	-0.343		
	(0.201)		
Constant	-77.309**	-75.622*	-40.368
	(29.619)	(30.394)	(39.197)
Observations	6,090	6,090	6,090
R-squared	0.61	0.59	0.31
Adjusted R-squared	0.61	0.59	0.31