Age- and Sex-Specific Estimates of Migration Flows in Asia-Pacific

Tianyu Shen, James Raymer, and Qing Guan
School of Demography, Australian National University

Version 10 (2021/11/22)

Abstract

Asia-Pacific is a diverse and large population region. Information on the flows of international migration within the region, including their age and sex compositions, is important for researchers and policymakers to understand the impact of migration on population change, employment, education and social developments. However, most countries in this region do not gather or produce flow statistics, and where they exist, there are differences in measurement and data collection procedures resulting in incomparable data. This study extends previous work that estimated the aggregate country-to-country flows for 53 Asia-Pacific populations to include age and sex characteristics. Apart from reported data from a few developed countries in the region, such as Australia and South Korea, age and sex information are inferred from the migrant stock data obtained from the United Nations' Population Division, flows of foreigners from Organisation for Economic Cooperation and Development (OECD) International Migration database, and estimated five-year migration flows in prior research. To estimate the age and sex patterns for each migration corridor, we use a log-linear model to combine the estimates of aggregate origin-destination flows with estimates of the age-sex profiles. The resulting origin-destination-age-sex flows of migration are then presented and assessed. This paper provides a basis for understanding the age and sex dynamics of migration in the Asia-Pacific region.

Key words: International migration, age, sex, Asia-Pacific, log-linear model

Correspondence to: tianyu.shen@anu.edu.au

1. Introduction

International migration is critical for understanding demographic change in the Asia-Pacific region. This region contains more than three-fifths of the world's population. There is ample evidence that migration intensities have increased in recent decades (e.g., Charles-Edwards et al. 2016; De Haas et al. 2020), despite only limited data are available on the actual annual international migration flows. Methods have been developed by researchers to estimate international migration flows in situations with poor-quality reported data. However, hardly any have focused on estimating annual migration flows in the Asia-Pacific (e.g. Raymer et al. 2019), and no estimates exist for the corresponding age and sex patterns. Previous efforts have estimated annual international migration flows for European countries with considerably more data (e.g., Raymer et al. 2011, 2013; Wiśniowski et al 2013, 2016; De Beer et al. 2010). Recently, Abel (2018) developed a methodology to estimate five-year international migration transition data by sex for most countries in the world based on international migrant stock.

Understanding the age and sex composition of migration flows is essential for researchers and policymakers to understand the impact of migration on employment, education, and social development. In this paper, we present a methodology to extend origin-destination-specific estimates of international migration flows in the Asia-Pacific region (Raymer et al. 2020) by age and sex. This includes flows amongst 53 Asia-Pacific populations and four macro world regions: Rest of Asia, Europe, Africa, and South and Central America. Specifically, we develop a statistical model to estimate annual migration flows by origin, destination, age and sex (ODAS) for the 53 Asia-Pacific populations using a multiplicative component framework. The goal is not to estimate the exact value as the "truth" is unknown to anyone, but to provide a basis for understanding the age and sex dynamics of migration in the Asia-Pacific region.

2. Background

Migration takes place for various reasons, such as education, employment and family reunion. Moreover there are people moved forcefully like human trafficking (De Haas et al. 2020). However, regardless of reasons or purposes, the basic data of the level and direction of the migration are largely missing (Willekens et al. 2016). There are different sources of migration statistics such as population registers or census, border control and surveys. These different sources have their own strengths and weaknesses in reflecting the migration flows. For population register or census, it is good to know the migrant stock inside the countries, but it does not have the same focus in terms of the flows of the migrants. As for the border statistics, it is a great way to capture the demographics of the arrivals if the countries have strong border control, but not for departure and irregular migrant. Moreover, many countries in Asia-Pacific might not have a strong border control as they are connected to on another geographically and human trafficking or forced labour is quite prevalent in this region (ILO 2017). Survey is a great way to gathering information on immigrant populations and their characteristics with extensive questions, but migrant population are usually underrepresented if it is a general purpose survey. Therefore, it is hard to know the flow of migrant between countries and even harder to know the basic demographic distribution (i.e., age and sex).

There are a few previous studies estimating flow with Asia-Pacific countries and most of them rely on the conversion of migrant stock data. From these estimates, we can get a sense of the level and direction of the migration. The first study is by Abel (2013; 2018), where he used demographic accounting method to distribute the migrant stock controlling for births and deaths in the destination as well as the distance between origin and destination. These estimates are the minimum migrant transitions to achieve the new migrant stock over five years or ten years. With this method he was able to convert the stock into bilateral flows by sex, but the age information was not included. Dennett (2016) proposed a slightly simpler method with two main estimations:

migration probabilities relative to the total stock population and total global migrant flows. With these two elements multiplied by each other, he calculated the bilateral flows. Another set of estimates by Azose & Raftery (2019) also used migrant stock with a pseudo-Bayesian approach. They modified Abel's (2018) method by adding an independent log-linear model, without diagonal interaction terms and producing a weighted average with Abel's estimates. A recent study by Raymer et al. (2020) adopted a very different approach called the generation-distribution model. Their model first estimates the emigration total with a set of covariates, such as population size and GDP for each origin, and then distributes the emigration into each destination using immigrant population sizes and bilateral trade.

Among these estimates, only Abel provided the origin-destination flows by sex. However, since the other two estimates are transforming the same sex-specific migrant stock data, it should be relatively straightforward to treat male and female migrants separately and produce the estimates by sex. On the contrary, the study by Raymer and the others (2020) would require extra steps to include sex estimation because the method used in either generation or distribution cannot distinguish female from male easily.

Efforts to estimate age- and sex-specific migrant flows are seen in European populations since there are more available data. For example, Raymer et al. (2011) used a multiplicative approach to incorporate an incomplete OAS (emigration by origins, age and sex) table and an incomplete DAS (immigration by destinations, age and sex) table with the estimated origin-destination (OD) flows for the 31 European countries. Similarly, Wiśniowski et al. (2016) disaggregated the OD flows by age and sex by the same multiplicative model with the Bayesian approach. The main difference between this paper and their studies is the availability of data. In Europe, although not every country reports the same amount of migration statistics and the numbers are often inconsistent, there are still considerable data to gauge the migration. Nevertheless, the data is way too scarce in Asia-Pacific countries.

Table 1. Sex and age ratios by country and region, 2010-2015

Region	Code / Country or area	UN stock sex ratio	UN Stock age ratio	Abel's flow sex ratio (origin)	Abel's flow sex ratio (destination)
E Asia	CHN China*	1.47	3.13	0.87	2.17
	HKG Hong Kong	0.70	0.99	0.41	0.69
	MAC Macao	0.85	1.79	0.46	0.89
	PRK DPR Korea	0.98	2.68	1.04	1.06
	JPN Japan*	0.84	3.92	0.04	0.95
	MNG Mongolia*	2.84	4.80	0.63	3.80
	KOR Rep. Korea*	1.27	5.54	1.46	1.17
	TWN Taiwan	NA	NA	NA	NA
S-E Asia	BRN Brunei Dar.	1.29	3.68	1.67	1.40
	KHM Cambodia	1.12	4.97	1.88	1.07
	IDN Indonesia	1.35	5.23	1.01	1.41
	LAO Laos*	1.57	4.94	1.01	5.06
	MYS Malaysia*	1.55	12.85	1.48	1.09
	MMR Myanmar*	1.19	3.41	1.02	0.03
	PHL Philippines*	1.07	2.97	1.00	0.98
	SGP Singapore	0.79	2.42	0.50	0.94
	THA Thailand	1.02	6.68	0.53	0.86
	TLS Timor-Leste	1.02	6.72	1.00	NA
	VNM Viet Nam*	1.40	4.94	1.12	0.99
S Asia	AFG Afghanistan	1.30	5.21	0.98	2.09
	BGD Bangladesh	1.12	5.17	1.14	∞
	BTN Bhutan	4.30	7.25	1.95	2.96
	IND India	0.94	1.14	1.12	0.66
	IRN Iran*	1.21	9.77	1.84	0.11
	MDV Maldives*	3.22	4.97	1.28	1.03
	NPL Nepal	0.49	4.49	3.12	0.00
	PAK Pakistan	1.11	0.99	1.23	0.68
	LKA Sri Lanka	1.13	3.97	1.96	0.58

Table 1 (Continued). Sex and age ratios by country and region, 2010-2015

Region	Code / Country or area	UN stock sex ratio	UN Stock age ratio	Abel's flow sex ratio (origin)	Abel's flow sex ratio (destination)
Oceania	AUS Australia	0.99	1.40	0.77	0.91
	NZL New Zealand	0.94	1.99	1.16	0.78
	ASM American Samoa	1.06	3.64	NA	NA
	COK Cook Islands	1.01	5.77	NA	NA
	FJI Fiji	1.15	2.99	0.99	0.94
	PYF French Polynesia	1.33	2.10	1.23	3.41
	GUM Guam	1.07	3.49	1.08	1.08
	KIR Kiribati	1.12	8.31	1.16	NA
	MHL Marshall Islands	1.59	5.49	NA	NA
	FSM Micronesia	1.15	3.77	1.00	NA
	NRU Nauru*	1.16	3.69	NA	NA
	NCL New Caledonia	1.16	1.90	1.72	1.03
	NIU Niue	1.18	5.01	NA	NA
	MNP N Mariana Islands	0.73	3.45	NA	NA
	PLW Palau	1.41	5.47	NA	NA
	PNG Papua New Guinea*	1.54	3.52	1.16	1.41
	WSM Samoa	1.03	6.28	1.01	0.48
	SLB Solomon Islands	1.27	3.21	1.00	NA
	TKL Tokelau	0.93	7.68	NA	NA
	TON Tonga	1.19	5.83	0.99	1.01
	TUV Tuvalu*	1.22	3.68	NA	NA
	VUT Vanuatu	0.99	5.03	∞	0.48
	WLF Wallis + Futuna	1.02	4.96	NA	NA
N America	CAN Canada	0.92	1.33	0.70	0.88
	USA United States	0.96	2.19	0.57	0.99
Other	ROA Rest of Asia	NA	NA	1.06	1.51
	AFR Africa	NA	NA	1.21	1.09
	EUR Europe	NA	NA	1.10	0.97
	SCA S+C America	NA	NA	1.04	1.06

Sources: United Nations (2019a, b), Abel (2018)

Note: Country or area with * are migrant stock by citizenship, others are by birthplace. Sex ratio is the number of male migrants to female; age ratio is the number of migrant under 50 to above 50. For Abel's estimates, we use the option with stock UN (2015), population UN (2015) and interval 5.

In this study, we extend on the previous estimates of 53 Asia-Pacific countries and areas (Raymers et al. 2020), where, countries in the Central and West Asia are excluded but the US and Canada are included because of their geographic locations and importance to migration connections with Asia-Pacific populations. The list of countries or areas is shown in Table 1. There are four broad world regions outside the Asia-Pacific region in the previous estimates: the Rest of Asia (ROA), Europe (EUR), Africa (AFR), and South and Central America (SCA).

3. Data

As identified in many prior studies, data paucity and inconsistency are the main challenges for research in migration and its impact (Abel & Sander 2014). Therefore, most data we can draw on in this paper are estimated values paired up with limited information from a few countries' statistical offices. The first source of data is a set of estimated origin-destination flows representing annual migration amongst 53 Asia-Pacific populations and four macro world regions between 2000 and 2019 (Raymer et al. 2020). The methodology underlying the estimates is divided into two parts. Emigration flows are first estimated for each population using a regression model based on key variables related to migration and the European data (Raymer et al. 2013). Since the European estimates have posterior distribution, the result includes uncertainty. These flows are then distributed across potential destination countries based on bilateral trade and immigrant stock statistics. Strong correlations between emigration and immigration are imposed at the end in line with migration theory. The OD estimates include 1,000 iterations for each corridor.

The second primary data is the International Migrant Stock 2020 from International Migrant Stock (UN 2020), which collects and estimates age and sex-specific migrant population stock for 52 destinations in Asia-Pacific excluding Taiwan every five years from

1990-1995 to 2015-2020. For the majority of countries in Asia-Pacific, stock data are the only source with age and sex patterns. Migrant stocks in most countries are defined by birthplace, while 14 out of 52 destinations in Asia-Pacific reported their migrant stock by citizenship. We treat the two definitions the same because we lack the resource and data to distinguish and reconcile the differences. Figure 1 presents the migrant stock in percentage by age and sex for nine selected countries (for demonstration). The sex ratios and age ratios for all countries and regions are shown in Table 1.

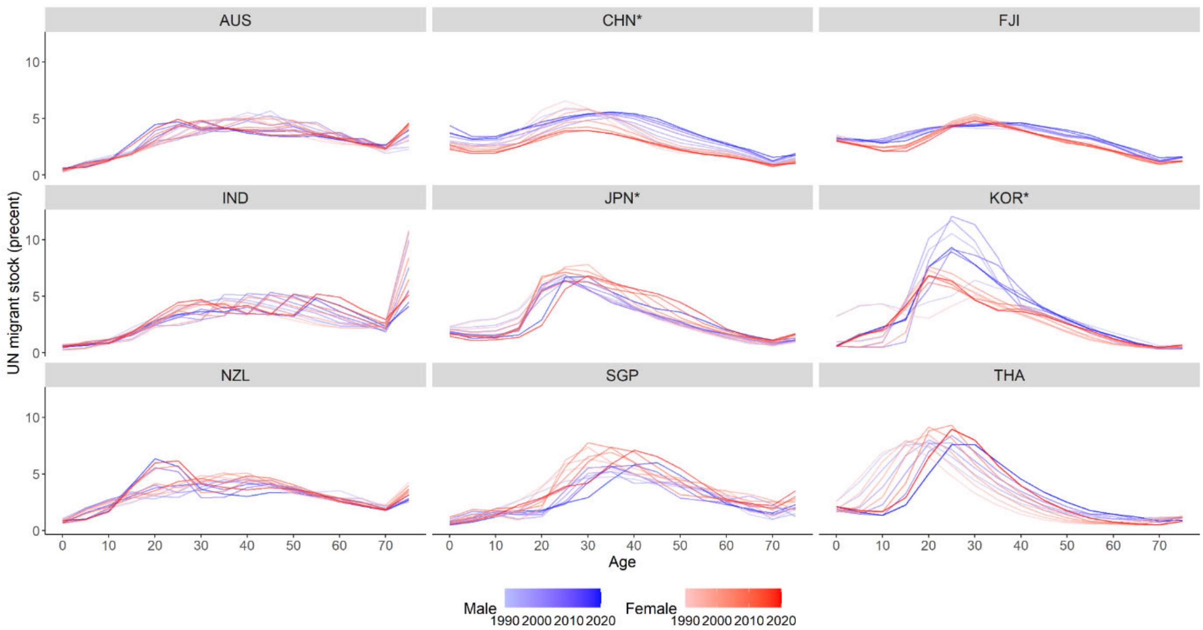


Figure 1. UN migrant stock by age and sex in percentage for selected destination, 1990-2020

Note: Country code with astride represents the definition of citizenship for migrants.

This paper also borrows from Abel (2018) estimates to enhance the estimation on sex-specific flows when no other information is available. Five-year transitions with the latest demographic and migrant stock data from UN are adopted as auxiliary information in our models. Table 1 presents the sex ratio of male to female by destinations and origins in Abel's estimate for 2010-2015. Most of the sex ratios are around one, with some extreme values such as Japan and Nepal, as well as missing values in Oceania countries.

Finally, reported data from OECD countries are used as inputs for baseline age-sex profiles. Australian, Korean and New Zealand's statistics offices (ABS 2020; Statistics New Zealand 2020; Statistics Korea 2020) publish the numbers of long-term immigrants and emigrants by age and sex from 2000 to at least 2017. Statistics Canada (2020) also publishes annual numbers of immigrants and emigrants by sex. OECD database includes some information on the origin of foreign immigrants or the destination of foreign emigrants by sex. However, the definition of foreign immigrants can also vary for some countries defining by birthplace and others citizenship or residency. We are also aware that there are other reported data. For instance, the ILMS database has sparse statistics on the sex of labour migrants in ASEAN (Association of Southeast Asian Nations) countries. Nevertheless, the major limitation is the scope of the migrants since labour migrants may not represent the age-sex structure of all the other types of migrants. Consequently, we use them to validate model estimates rather than model inputs.

4. Methodology

This paper adopts Raymer et al.'s (2011) method to disaggregate origin-destination flows by age and sex. Their method starts from a complete OD (Origin by Destination) two-way table of migration flow with two main effects, overall proportion from each origin (O_i) & overall proportion to each destination (D_j), and one interaction, (OD_{ij}), $n_{ij} = (T)(O_i)(D_j)(OD_{ij})$, where n_{ij} is the observed flow of migration from origin *i* to destination *j* and *(T)* is the overall migration level. The two-way interaction (OD_{ij}) represent the ratio of observed flow to the flow determined solely by (O_i) & (D_j). Age and sex information are then incorporated by log-linear model into an ODAS four-way (origin by destination by age by sex) table via offsets. Here are the multiplicative components of a saturated ODAS table,

$$n_{ijxy} = (T)(O_i)(D_j)(A_x)(S_y)(OD_{ij})(OA_{ix})(OS_{iy})(DA_{jx})(DS_{jy})(AS_{xy})$$
$$(OAS_{ixy})(DAS_{jxy})(ODA_{ijx})(ODS_{ijy})(ODAS_{ijxy}).$$

Similar to Raymer et al.'s (2011) approach, we assume that recently produced origin-destination flow estimates from Raymer et al. (2020) are true and then add in age and sex main effects and two-way or higher dimensions interactions. European countries have considerably more data regarding the age and sex pattern of emigration and immigration. In Asia-Pacific however, migration data are extremely sparse and hardly include any age- and sex-specific details. Therefore, the aim of this paper is to extend previous work by including age and sex characteristics into country-to-country migration flows estimates for 53 Asia-Pacific populations. There are four stages of the model as shown in Table 2, and each stage generates a set of stand-alone estimates that can be assessed. Although they are stand-alone stages, the design of the framework is to enhance the estimates by adding more or modifying effects using auxiliary information. This section only presents a brief overview of the method. For technical details, refer to Appendix 1.

Stage		Aim		Data	Main procedure	Existing effects	New/ revised effects
1		Integrate overall age and sex profile		I. Estimates of migrant flows by origin & destination (2000-2019) II. Australia immigrant & emigrant by age & sex (2000-2019) III. South Korea immigrant & emigrant by age &sex (2000-2018) IV. New Zealand immigrant & emigrant by age & sex (2001-2019)	1. Generate general age profiles by sex from reported values (Data II, III, and IV) with loess regression. 2. Set the overall sex ratio to one. 3. Integrate the age and sex profile with Data I.	(O_i), (D_j), (OD_{ij})	(A_x), (S_y), $(\boldsymbol{AS_{xy}})$
2	2a	Integrate age and sex profile by destination	Estimate foreign immigrants age and sex profile by destination	I. UN International Migrant Stock 2020 by age and sex for 52 destinations (1990, 1995, 2000, 2005, 2010, 2015, 2020) II. UN World Population Prospect 2019 Life Table by sex for 52 populations (1990, 1995, 2000, 2005, 2010, 2015, 2020)	1. Compute the change in stock by reverse cohort-component projection accounting for survival. 2. Rescale the age and sex pattern to positive and extend older age groups (75-79, 80-84, and 85+) with model migration schedule.	(O_i), (D_j), (A_x), (S_y), (AS_{xy}), (OD_{ij})	(S_y), (AS_{xy}), (DS_{jy}), (DA_{jx}), $(\boldsymbol{DAS_{jxy}})$
	2b		Estimate returning nationals age and sex profile by destination	I. Australian-born arrival by age and sex (2000-2013) II. South Korean national arrival by age and sex (2000-2018) III. OECD outflow of foreigners by sex and destination from Australia (2000-2016), South Korea (2000-2018) and New Zealand (2000-2018)	1. Generate general age profiles by sex for returning nationals from reported values (Data I and II) with loess regression. 2. Adjust sex profile by destination with Data III.		
	2c		Estimate the proportion of returning nationals	I. Estimates from Stage 1 (2000-2019) II. Net migration rate from estimates of migrant flows by origin & destination (2000-2019) III. Migration efficiency from estimates of migrant flows by origin & destination (2000-2019) IV. Percentage of migrant population calculated from International Migrant Stock 2019 & UN World Population Prospect 2019 V. Australia immigrant by place of birth (2000-2013) VI. South Korea immigrant by citizenship (2000-2018) VII. New Zealand Immigrant by prior residency (2001-2019)	1. Estimate logistic relationships between reported proportions (Data V, VI, and VII) and variables related to migration impact (Data II, III, and IV). 2. Apply the logistic relationships to all destinations with Data II, III, and IV and obtain the average proportion of returning nationals for each destination. 3. Join the two profiles above by this proportion for each destination. 4. Integrate the age and sex and destination profiles with Data I.		

Table 2. A non-technical summary of methodology (cont.)

Stage		Aim		Data	Main procedure	Existing effects	New/ revised effects
3	3a	Adjust sex profile by origin and by destination	Adjust sex effect for four rest of the world (ROWs) destinations	I. Estimates from Stage 2 (2000-2019) II. Abel's Estimates of migration transition by destination & sex (2000, 2005, 2010)	1. Calculate the average sex profile for the four ROWs as destinations in Abel's estimate (Data II) across time. 2. Combine sex profiles by destination of Stage 2 (Data I) and Abel's average by weight. 3. Obtain new age and sex and destination profiles including the new overall sex profile.	(O_i), (D_j), (A_x), (S_y), (AS_{xy}), (OD_{ij}), (DS_{jy}), (DA_{jx}), (DAS_{jxy})	(S_y), (AS_{xy}), $(\boldsymbol{DS_{jy}})$, (DAS_{jxy})
	3b		Adjust sex effect for origins	I. Estimates from Stage 2 (2000-2019) II. Abel's Estimates of migration transition by origin & sex (2000, 2005, 2010)	1. Calculate the average sex profile for each origin in Abel's estimate (Data II) across time. 2. Combine sex profiles by origin of Stage 2 (Data I) and Abel's average by weight controlling for the new overall sex profile in Stage 3a. 3. Integrate these new profiles with Data I.	(O_i), (D_j), (A_x), (S_y), (AS_{xy}), (OD_{ij}), (DS_{jy}), (DA_{jx}), (DAS_{jxy})	$(\boldsymbol{OS_{iy}})$
4		Replace profiles by reported data		I. Estimates from Stage 3 (2000-2019) II. Australia immigrant & emigrant by age & sex (2000-2019) III. South Korea immigrant & emigrant by age &sex (2000-2018) IV. New Zealand immigrant & emigrant by age & sex (2001-2019) V. Canada immigrant & emigrant by sex (2000-2019)	Replace the estimates with reported age and sex profile controlling for OD and AS	(O_i), (D_j), (A_x), (S_y), (AS_{xy}), (OD_{ij}), (DS_{jy}), (DA_{jx}), (OS_{iy}), (DAS_{jxy})	(DS_{jy}), (DA_{jx}), (OS_{iy}), (OA_{ix}), (DAS_{jxy}), (OAS_{ixy})

Table 2. A non-technical summary of methodology (continue)

Noted that the uncertainty mentioned in this paper cannot be interpreted as the traditional prediction interval and it is more of a sensitivity range based on the variation of the data. We simulate the age and sex profile for 1000 times to match it with the number of iteration in flows by origin and destination. Therefore, the uncertainty in the paper also incorporated the uncertainty from the age-sex-aggregated estimates.

Stage 2 aims to incorporate immigrants' age and sex profile for each destination. Migrants arriving in a country can be native-born (nationals) or foreign-born (foreigners) with the age and sex distribution for these two groups likely being different. For example, about 40% of foreigners arrived in Korea between 2000 and 2018 were female. On the contrary, more than 50% of nationals returned to Korea were female. Therefore, we estimate separate destination by age by sex (DAS) tables for the native-born (nationals) and foreign-born (foreigner) populations and combine them using weights to obtain the full DAS table for each destination.

In Stage 2a, we first calculate the DAS table for foreign-born (foreigners) from the UN international migrant stock. Although the migrant stock has an age and sex profile, it is not feasible to apply the age and sex schedule to the migration flow table directly. For countries, shown in Figure 1, with the majority of the migrant stocks being short-term foreign labour, the stock data might seem to have a plausible migration age structure peaked at working age. However, it is not perfect especially for major migration destination countries, such as Australia, where older populations are over-represented because of the accumulation of stock. Moreover, there are other countries in Southern Asia and Oceania that have peculiar shapes (e.g., peak at children age groups and transitions of age and sex profiles over time).

Therefore, we transform these data into changes in stock to simulate the flow statistics. The changes in stock are projected with a reverse cohort component method to backcast the stock five years ago adjusting for the survivorship. Then the difference between the back projection value and the actual stock is the change in stock. It is based on the assumption that no migrant exits the population, so this change in stock is a minimum transition between two time points of two subsequent age groups. Since we only know this transition takes place in either age group, we arbitrarily assign a normal distribution of mean at 0.5 and standard deviation at 0.2 to divide the change in stock into two age groups randomly. Again, 1,000 iterations are performed where each draws a probability from the normal distribution.

The stock data only provide information until age 70, but our estimates have the open age group at 85 years old. We adopted part of the migration schedule proposed by Rogers and Castro (1981) for the labour force decent[1] , and apply an exponential distribution for age groups over 30 (some countries are 25 or 35 depending on the sum of square error between the smooth curve and the original). The smooth curves then give predictive values for age groups from 30-34 to 85+.

After obtaining the age and sex distribution of foreign-born (or foreigner) arrival, we estimate the age and sex profiles for the native-born (or national) returning. Only two countries provide detailed statistics by age and sex on immigration of the native-born (Australia) and nationals (Korea) (hereinafter referred to as "nationals"). The national naïve AS schedule is calculated using the same method in Stage 1 --- loess regression to smooth the age distributions of arriving nationals by sex. OECD has the outflow of foreigners by sex and destination from Australia, Korea and New Zealand. Assuming that the sex ratio of the outflow of foreigners is equal to the inflow of nationals of the destination, we derive the sex ratio of nationals arriving. The magnitude of flows from Australia, Korea and New Zealand differs, so we rescale them to the same weight and calculate the average sex ratio of return flow to each country. Only 80% of the data is randomly selected as a subsample when estimating naïve AS schedule and sex ratio national arrivals.

[1] a gradual decline after the migration peak at early adulthood (Rogers and Castro (1981))

With the proportion of foreigner and national arrivals by age and sex for each destination, Stage 2c is to combine the foreigner and national destination by age by sex (DAS) profile into one by weights and integrate the full profiles to Stage 1 estimates. To obtain the combined profile, we need the proportion of national arrivals among all arrivals (i.e., the weight). Intuitively, the major receiving countries would have more foreigners arriving, and the major sending countries would have higher national arrivals since most of the emigrants would return. However, there is no empirical study on this proportion. Therefore, we assume this weighing proportion follow logistic function and draw on the average relation between reported data (from Australia, South Korea and New Zealand) and three variables measuring the exchange of migrants and the impact of migration (i.e., net migration rate, the effectiveness of migration and logged proportion of migrant stock).

After the foregoing steps, we can add up these two DAS profiles by the weight of each country. Our new model now includes three more interaction terms, where three two-way interactions, $\left(AS^{*}_{xy}\right)$, $\left(DA_{jx}\right)$ and $\left(DS_{jy}\right)$ automatically derived from the DAS table. One issue with this log-linear model is that the sex component from $\left(DAS_{jxy}\right)$ differs from the previous sex main effect $\left(S_{y}\right)$. Since the previous sex ratio at one is purely naïve, we opt for the sex ratio (males to females) from DAS table and derive a new overall sex effect $\left(S^{*}_{y}\right)$, which is between 1.10 and 1.16 with a median at 1.13 in 2019. $\left(S_{y}\right)$ would hence fluctuate slightly each year due to the uncertainty. On the other hand, the age main effect (A_{x}) is kept as it is based on the reported data.

Stage 3 is to add in OS interactions (OS_{iy}) to all 57 populations and DS interactions $\left(DS_{jy}\right)$ for the four Rest of the World regions from Abel's estimates (2018) to enhance our results. In Abel's study, migration flows by origin, destination and sex are estimated for most countries in the Asia-Pacific. However, Abel's estimates include many zeros in origin by destination flows and 30 percent of non-zero flows are highly skewed regarding sex (single sex accounting for more than 80 percent of the flow). Due to these limitations, we only borrow Abel's (OS_{iy}) and (DS_{jy}), rather than the detailed three-way interaction (ODS_{ijy}).

In Stage 3a, we adjust the sex ratios of four rest of the world (ROW) destinations, namely Africa, the rest of Asia, Europe and South and Central America. The sex ratios of immigrants to these destinations were set at one in Stage 2, which overlooks the effect of Middle East attracting substantial numbers of male manual labour. Therefore, we combine Abel's (DS_{jy}) and (DS_{jy}) from Stage 2 by a proportional weighting to generate a new proportion for these four ROWs. The proportional weighting is assumed a normal distribution with the mean at 0.5 and standard deviation at 0.2. Leaning to a more conservative approach, we choose not to apply Abel's (DS_{jy}) directly due to more extreme values in that estimates. Since (DS_{jy}) of other destinations is fixed, the sex main effect (S_y^{**}) increases slightly to a median at 1.16 in 2019, which is used to control for (OS_{iy}) in Stage 3b.

Stage 3b incorporate (OS_{iy}) into the estimates by adopting Abel's (OS_{iy}). Figure 2 shows the proportion of female emigrants by origin in 2019 with 80 percent prediction intervals (from 10 percentile to 90 percentile). Triangles represent the average of Abel's (2018) estimates over time, and error bars are Stage 2 and 3 results in black and red respectively. Since results from Stage 2 do not include assumptions in OS interaction, black error bars are merely artefacts of the interactions of (DAS_{jxy}) and (OD_{ij}). After combining (OS_{iy}) from Stage 2 estimates with Abel's using the same normal distribution as weight as above to avoid extreme values, as well as controlling for the new sex main effect (S_y^{**}), we obtain red error bars and Stage 3 estimates.

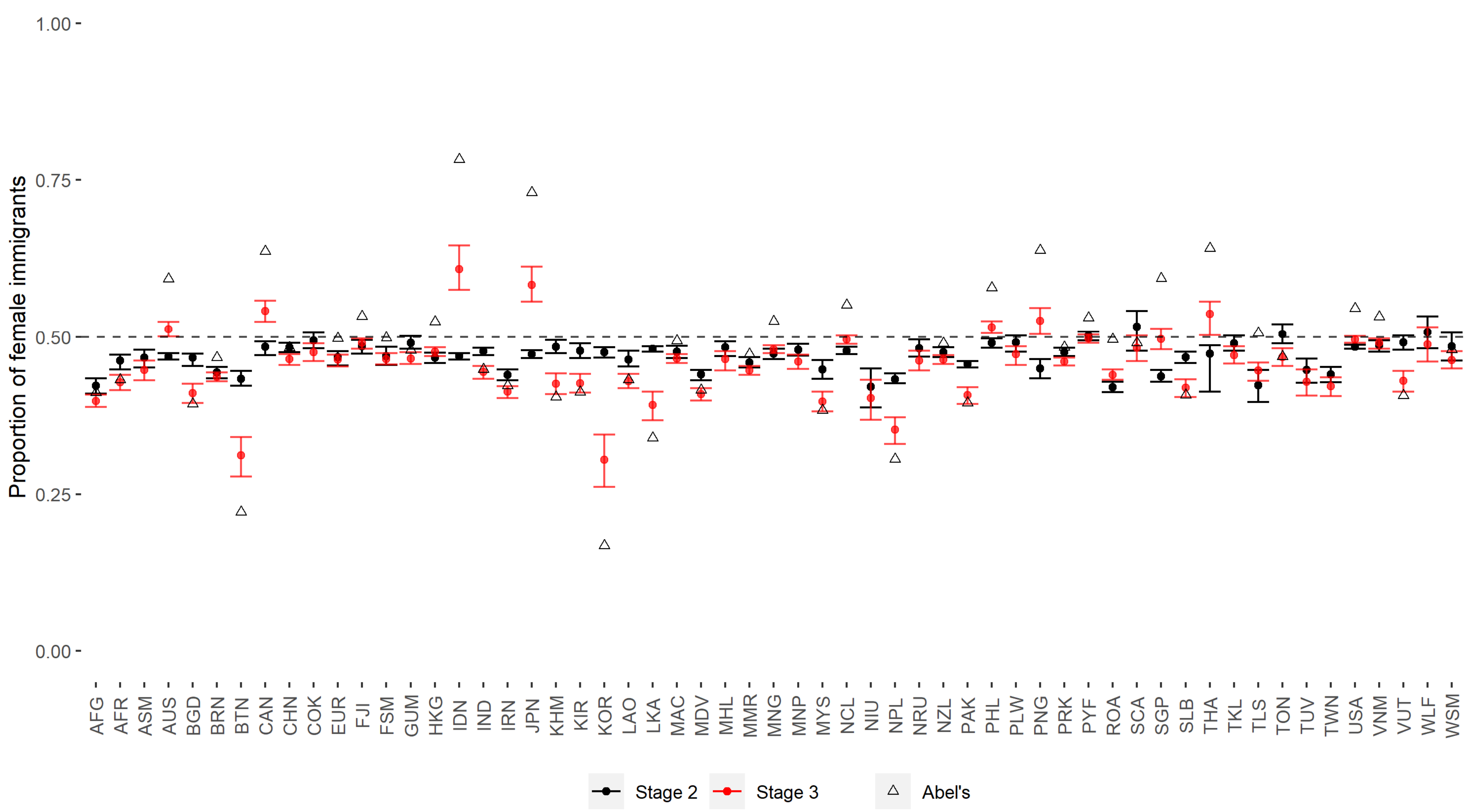


Figure 2. Proportion of female emigrants by origin and sex, 2019

We compare the result from Stage 3 with the relatively good data in Figure 3 and 4. Australia, Korea and New Zealand have detail age and sex components for both emigrants and immigrants, while Canada only has the number of immigrants and emigrants by sex. The reported value is exactly for 2018 (when all four countries have available data) and Stage 3 estimates are values of that year with 80 percent prediction intervals, however Abel's estimates are the average of three point estimates in 2000, 2005 and 2010. Although these data are reliable, none of their definitions perfectly align with the 1998 United Nations recommendations. For more definitions see Appendix 1.

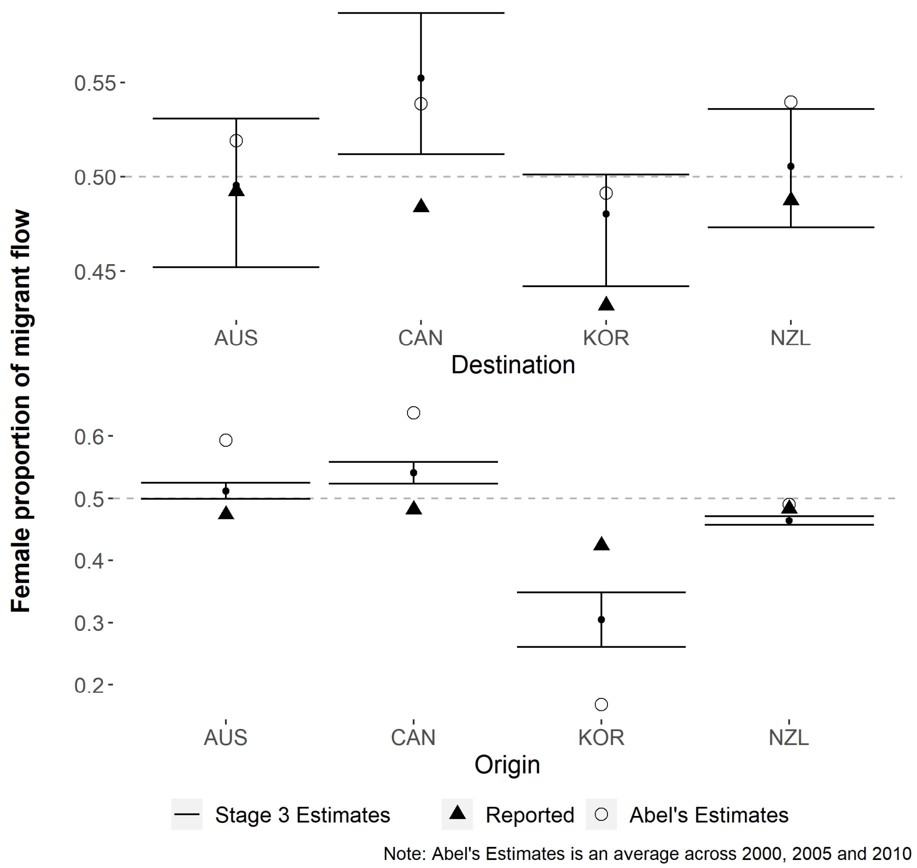


Figure 3 Comparison between estimated and reported of immigrants and emigrants by the female proportion, 2018.

Since Stage 3 incorporates Abel's estimates of migrant flows by sex, Figure 3 also includes Abel's results in circles. Figure 3 compares the sex component of Australia, Canada, Korea and New Zealand. For the destination, the estimated proportions of Canada in Stage 3 and Abel's study are on different sides of the reported. However, Stage 3 estimates are slightly closer to the reported data with the predictive interval. Regarding the origin, Stage 3 estimates are also closer to the reported, but Australian and Canadian results still flipped to the other side. Overall, the Stage 3 Estimates appears to perform better than Abel's estimates.

Figure 4 shows the estimated age and sex component for immigrants to and emigrants from Australia, Korea and New Zealand in 2018. Although the estimates do not follow the reported data perfectly, the peaks in particular, the overall trends however are mostly captured. Noted that the models produce the age and sex schedules based on very limited data and do not directly use any single country's flows as an input.

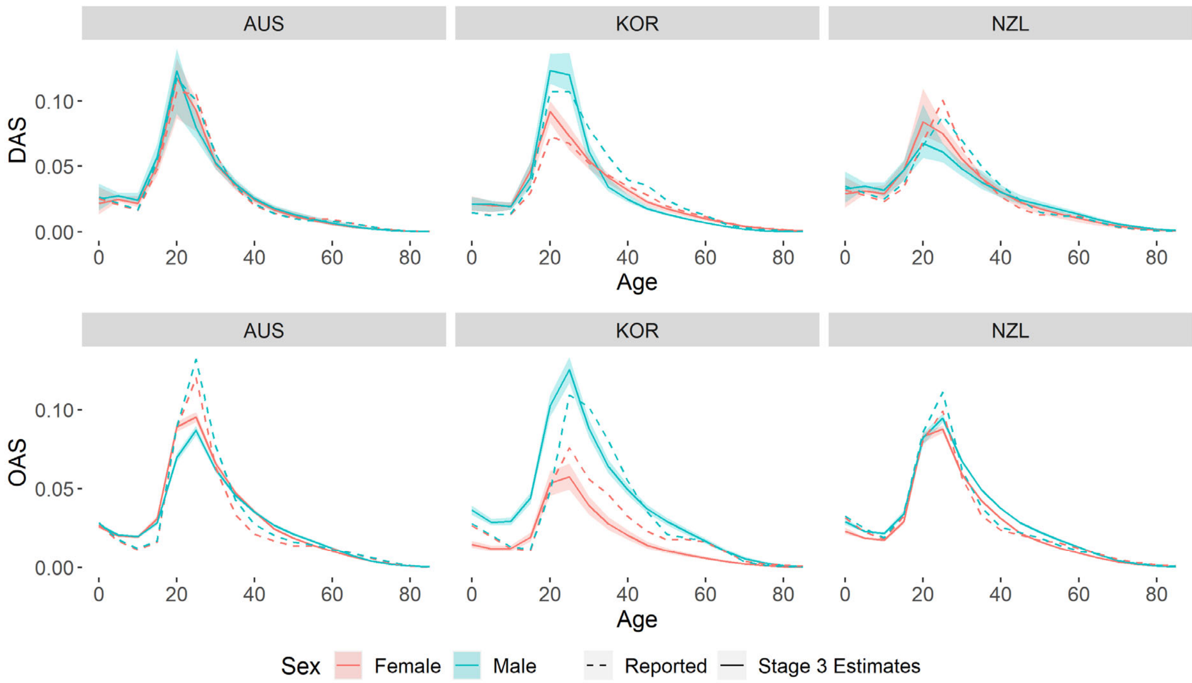


Figure 4 Comparison between estimated and reported immigrants and emigrants by the proportion of age and sex, 2018.

Therefore, in Stage 4, we replace the estimates with reported data for the four countries with relatively good international migration statistics by age and sex. Using IPF and maintaining the main sex and age effect, other countries should be adjusted accordingly. However, the effect on other countries is very limited, because of marginal adjustment compared to the overall magnitude.

5. Results

In this section, we present the Stage 4 estimates and validate them against a few reported data sources[2]. Three types of measurements are used to present the final estimates. The first one is the migration level, which is the direct output of the model. Comparing the level is straightforward to know which flows is the biggest, but it might bring difficulties to grasp the variation among age groups and sexes. The second measurement, proportion, focuses solely on the age and sex profile of the migrant. The proportion has been calculated multiple times in the method section, where each age and sex migration level is divided by the total level of that destination or origin, or corridor. However, this approach overlooks the population structure of the sending or receiving country.

The last set of measurements are normalized Age at Peak and Peak Migration Intensity (Bernard et al., 2014). Age at Peak describes the age at the highest migration intensity. Peak Migration Intensity (PMI) describes the level of migration at that peak age. To get these two measurements, age-specific immigration rates (ASIR) and age-specific emigration rates (ASER) are used. ASIR and ASER are calculated by the number of immigrants and emigrants

[2] One exception is the validation with Australian data by destination and origin, where 2013 results are used because the latest reported data available end in 2013.

respectively at each age group divided by the population at that age group[3]. The annual population by five-year age groups in each country can be obtained from UN WPP2019. However, population in some of the small Oceania countries are missing. Moreover, for the older age groups with many 0s, we randomly impute numbers from 100 to 499 to avoid undefined rates. To remove variations in the Asia-Pacific countries' overall migration intensities we rescale the sum of ASMR for each country to one. These three types of measurements have their strengths in showing certain perspectives of migration, so we adopt them all in this section to provide different lenses to the results.

Figure 5 presents the raw output of Stage 4 estimates --- the flows between different regions by sex in three age groups, namely 0-4, 25-29 and 60-64 in 2019. The scale is fixed for all circle plots at the maximum level among all age groups and sexes, so the width of each arrow can be compared across, which represents the level of the flow. The biggest flows are around the labour force peak --- age 25-29 in Figure 5, followed by age group between 0 and 4. The smallest flow here is at the older age group presented in the Figure. This distribution by age is consistent with age profile shown in Figure 4 and Figure A1.

The male and female flows in each age group appear to be largely identical, due to partly the sex ratio of all migrants close to one and partly the aggregation of countries in the sub-regions. However, there are visible differences between sexes in each age group. On the one hand, the level of the total inflows and outflows are fairly different, especially in the age group 25-29. From panel c and d, we can see that the size of female and male migrants from and to Oceania or Africa for example are about the same, while female migrants from and to Rest of Asia or Southern Asia are significantly less than their male counterparts. Literature also suggests high migration from Southern Asia (India in particular) to the Gulf States where man

[3] Note that ASIR would not have real meaning, as the denominator is not the population at risk of immigration. But they can still be used as a comparison index informing the relative size of immigration at different age groups.

worked in male-dominant occupations as contract worker or manual labour (Doreen, 2011; Castles et al., 2014). This kind of difference can be observed across age groups.

On the other hand, the relative sizes of flows to and from each region vary between sexes. The straightforward way to identify is from the order of the arrows. The arrows of each region are sorted by their direction (first sending then receiving) and width. In other words, the arrows closer to 0 on the scale represents a larger flow. For instance, look closely at Eastern Asia in panel c and d, both males and females migrate to Europe the most with comparable volume, but for female migrants the second biggest destination is not within Eastern Asia but Northern America. Turning to the receiving side, most female migrants are originated from Southeast Asia, followed by Eastern Asia and Southern Asia. Yet, males have the order of Eastern Asia, Southern Asia and Southeast Asia. Furthermore, these orders are not necessarily the same for each age group though they bear much resemblance.

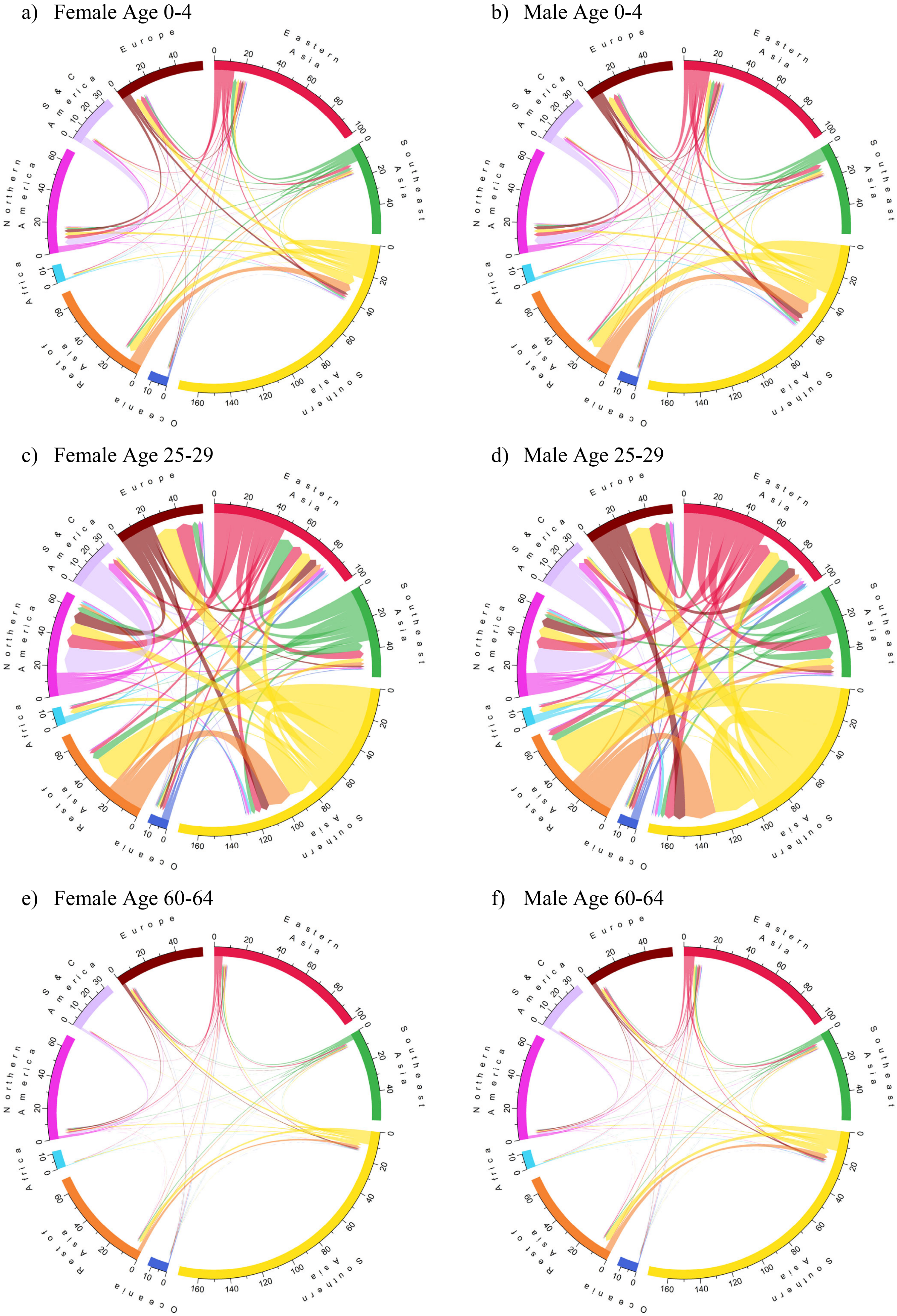


Figure 5. Stage 4 estimates for three age groups by sex aggregated to regions (ten thousand), 2019

Next, we examined the normalized migration intensity. The migration intensity by age and sex for each country is not presented, instead the information of the peak is extracted and compared. The PMI is the sum of normalized migration intensity of the age group at peak with the two adjacent age groups. The normalised migration intensity add up to one, so the higher the PMI, the more concentrated the age profile appears. As for the Age at Peak, it is estimated from the Age group at Peak and the PMI of each age group. We first assume the Age at Peak is at the mid-age of that age group (e.g., 22.5 for age group at 20-24). Then by comparing the PMI of the two adjacent age groups, we move towards the Age at Peak towards the upper or lower end of the age group proportionally. In other words, if the two adjacent age groups have about the same PMI, the Age at Peak would be estimated at the mid-age. When the older age group has much larger PMI, the Age at Peak would be close to the older bound of the Age group at Peak, and vice versa.

The two panels in Figure 6 present somewhat similar information with a shift in focus. Panel a focuses on correlation and uncertainty, where the age group at peak intensity is omitted, while panel b shows the age at peak for each sex and direction. From panel a, we can see a moderate positive correlation (0.53) between emigration intensity and immigration intensity. The major migration destinations, such as Japan, Hong Kong and Singapore appear to have more concentrated age profiles, especially for immigration. As for some sending countries like India, Afghanistan and Laos, they have more disperse shapes since their migrant types are likely more diverse.

From panel b, it is not hard to spot that the Age of Peak for female are generally younger than that of male. The average age of female immigrants is 26.5 and emigrant is 26.0, while male immigrants is 27.0 and emigrant is 27.2. For most countries, the Age of Peak centre around the average and the sizes of bubbles representing PMI are about the same for both sexes.

There are a few small countries having rather high Age of Peak likely because of their nonstandard age schedules.

a) Correlation of Peak Migration Intensity with a range of uncertainty

b) Age at Peak intensity and the peak intensity

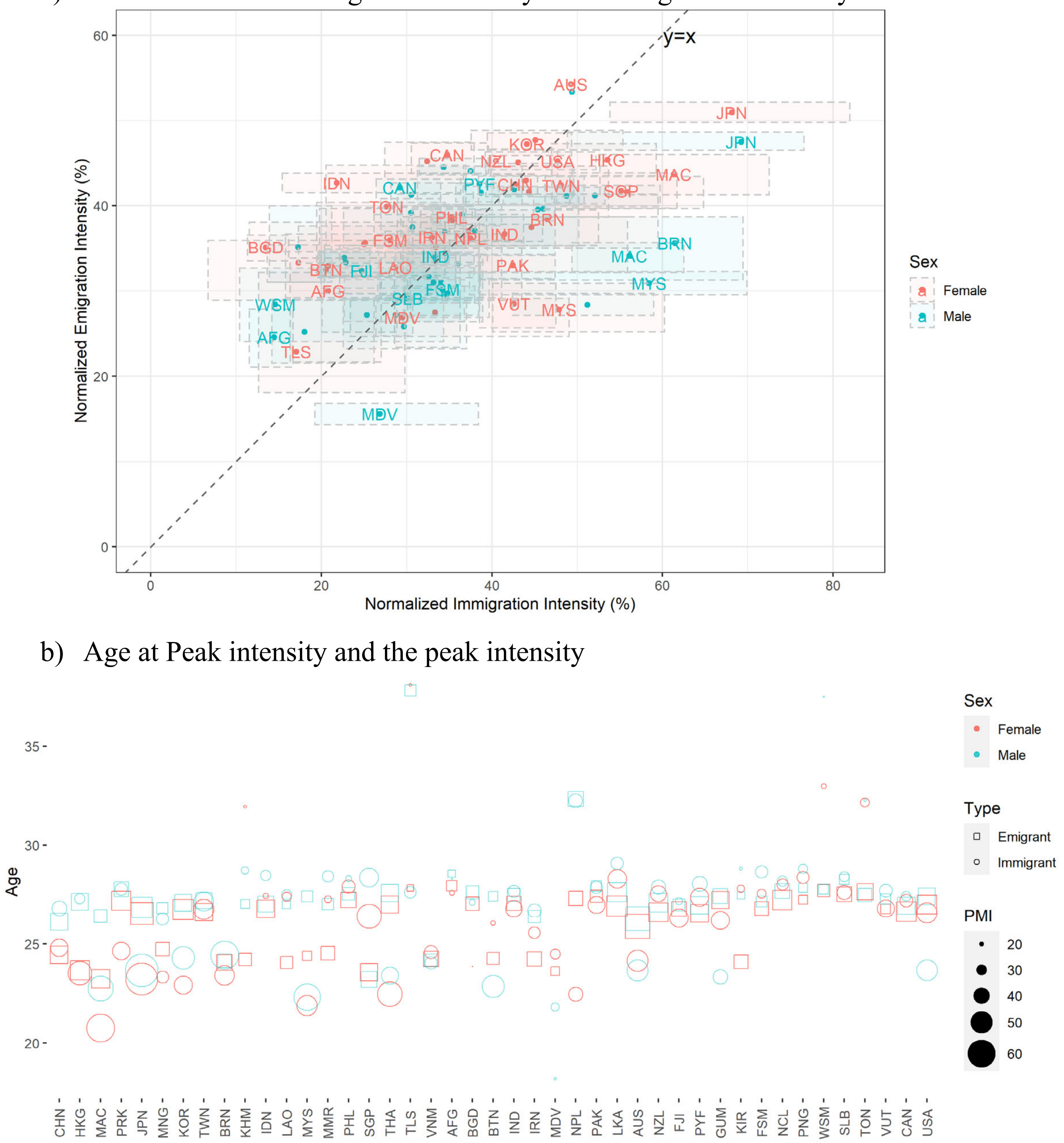


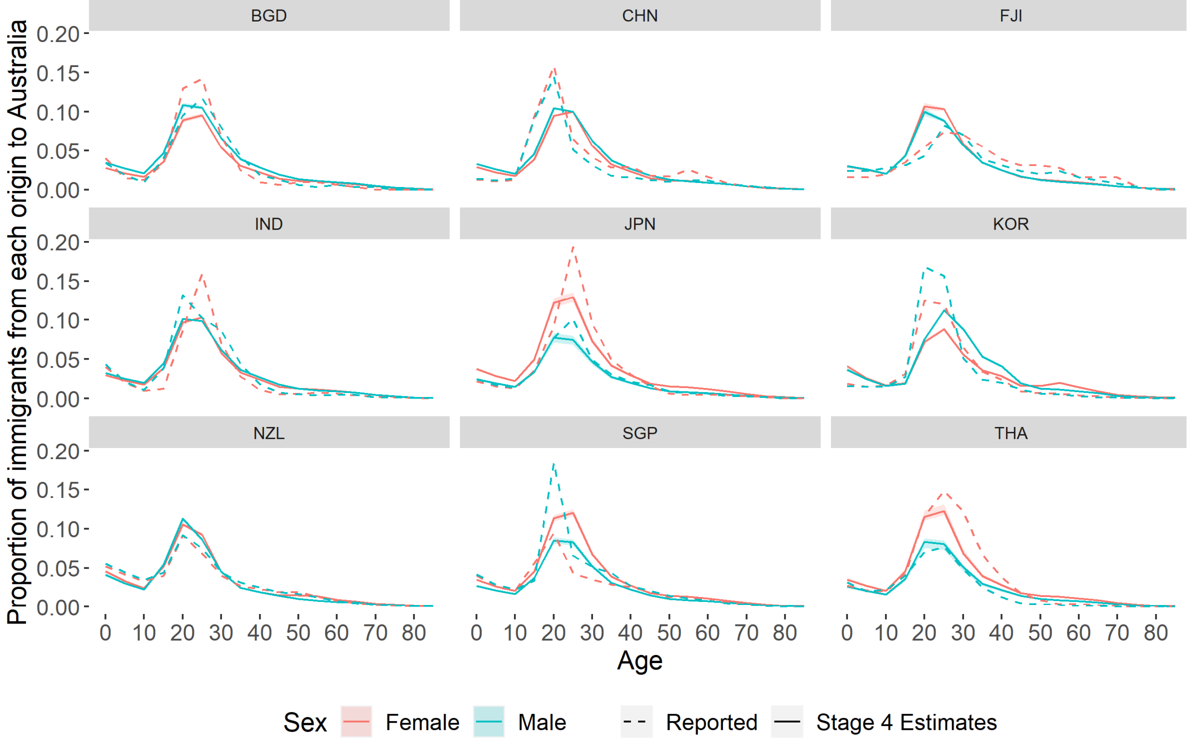
Figure 6. Peak Migration Intensity and Age at Peak in final model estimates, 2019

Figures 7 and 8 are validations against reported data. Figure 7 shows the age and sex proportions of immigrants to and emigrants from Australia for nine selected countries in 2013.

The dashed lines represent the reported data from ABS, and the solid lines are Stage 4 estimates with the 80 percent prediction interval. The uncertainties are relatively small compared to other countries because the total number of immigrants and emigrants by sex and age are replaced by reported values in Stage 4. Since we are only looking at Australia, the proportion of migrants is the simplest way to compare across migration corridors. The estimates roughly follow the pattern of the reported values, but they seem to have difficulty capturing the sharp peak and tend to be flatter shapes. This could be the effect of the overall age effect (A_x), which is a profile with a small tip at the first age group and a broad peak around age 20-30. However, the tip at the first age group in reported data is not always as high as the estimates, especially for emigration in panel b. As a result, the labour force peaks of the estimates are generally underestimated.

In Figure 8, the estimates are compared with the ILMS database, which primarily focuses on labour migration from ASEAN countries. Error bars are Stage 4 estimates in 2013 with 80 percent interval and the triangles are the average across all reported years in that countries. Most of their reported data are the total numbers of migrants without sex information. We can only find data from these six countries and only one of them (Indonesia) has information on both emigrants and immigrants. The model appears to underestimate the female proportion of migrants in Indonesia, Laos and the Philippines, while overestimating Myanmar and Thailand. For the underestimations, they are around 10 percent lower, but the overestimations seem to be quite big. It could be the problem with the reported values, which only capture the labour outflow and their female proportions of migrants are lower than 0.3.

a) immigration to Australia



b) emigration from Australia

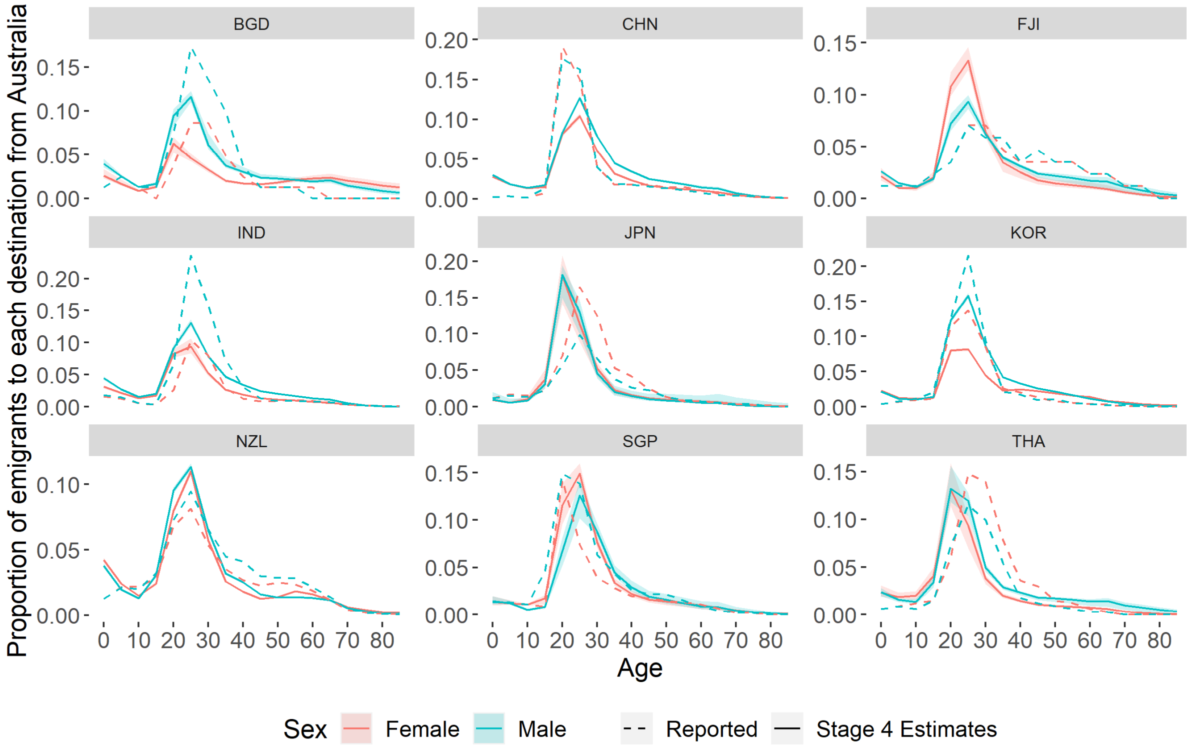


Figure 7. Comparing Stage 4 estimates in 2013 with reported age and sex proportion of migration from/to Australia

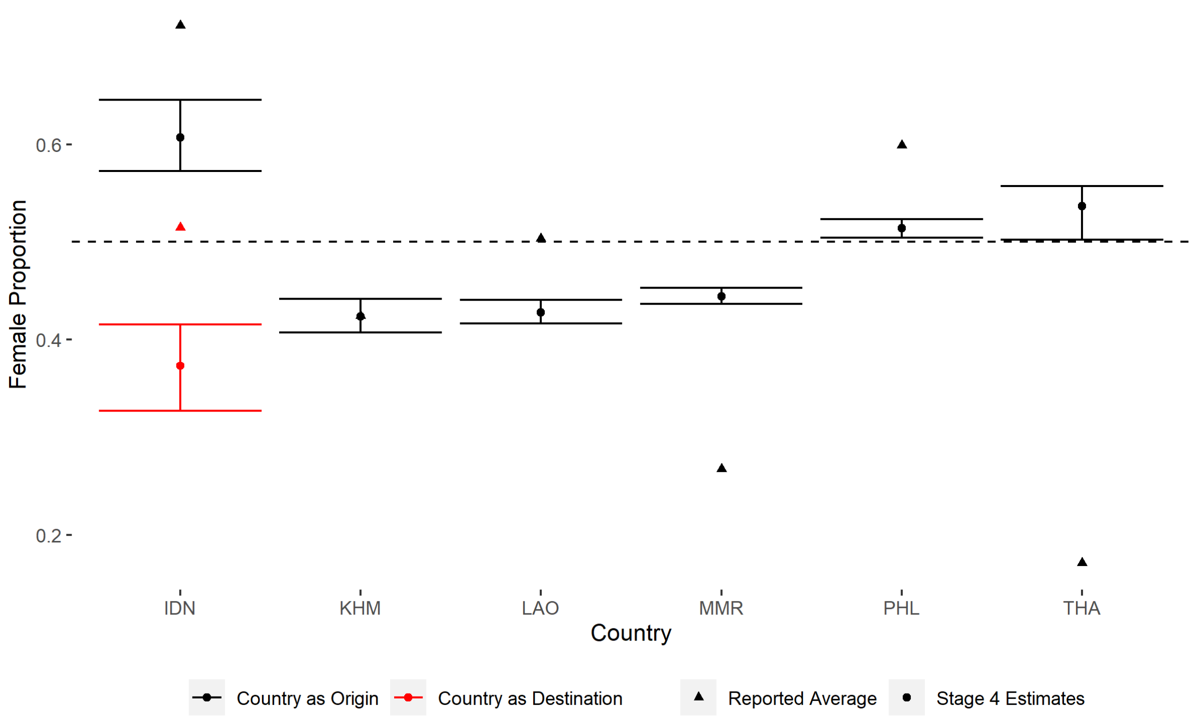


Figure 8. Stage 4 estimates against reported female labour migrant proportion in ILMS, 2013

6. Discussion

This paper provides a methodology to estimate the age- and sex-specific migration flows in Asia-Pacific countries in the context of very limited data available. The estimates are built within a multiplicative framework, which allows us to include borrowed information from different, likely incomparable, data sources relatively easily. The final set of estimates is built on a series of assumptions based on empirical findings found in other migration data situations.

With the multiplicative framework design, new information can be incorporated into the model to inform each multiplicative component effect. For further study, more interaction effects can be considered should new data emerge, such as trends in the components over time and more distinct age profiles by country of origin. The current results do not include time effects. Moreover, there may be some statistics from individual countries inaccessible publicly that could be potentially integrated into the estimates. Finally, further work is needed to validate the age- and sex-specific estimates. One way, for example, could be to insert the estimates into a cohort component population projection and compare the resulting population estimates with observed or reported figures. Another possible path should be to consult these results with country-level experts as to their plausibility.

In conclusion, this paper provides a basis for beginning to understand the age and sex dynamics of migration in the Asia-Pacific region. By doing so, we hope that more attention will be drawn towards understanding migration in this region and initiating further exploratory studies to determine the origin, destination, age and sex patterns of migration.

References

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

Abel, G. J., & Sander, N. (2014). Quantifying Global International Migration Flows. *Science, 343*(6178), 1520-1522. doi:doi:10.1126/science.1248676

Abel, G. J. (2018). Estimates of global bilateral migration flows by gender between 1960 and 2015. *International Migration Review*. doi:10.1111/imre.12327

Abel, G. J., & Cohen, J. E. (2019). Bilateral international migration flow estimates for 200 countries. *Scientific Data, 6*(1). doi:10.1038/s41597-019-0089-3

Azose, J. J., & Raftery, A. E. (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. doi:10.1073/pnas.1722334116

Bernard, A., Bell, M., & Charles-Edwards, E. (2014). Improved measures for the cross-national comparison of age profiles of internal migration. *Population Studies, 68*(2), 179-195. doi:10.1080/00324728.2014.890243

De Beer, J., Raymer, J., Van der Erf, R., & 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*(4), 459-481.

Dennett, A. (2016). Estimating an Annual Time Series of Global Migration Flows – An Alternative Methodology for Using Migrant Stock Data. In A. Wilson (Ed.), *Global Dynamics: approaches from complexity science* (pp. 125-142). Chichester, UK: John Wiley & Sons.

De Haas H., Castles S. and Miller M. J. (2020). *The age of migration: International population movements in the modern world, 6th Edition*. London: Red Globe Press.

Doreen, Storbeck (2011). Indian Labour Migration to the Arab Gulf States: The Impact of a Growing Interdependence. Internationales Asien Forum.International Quarterly for Asian Studies, 42(1), 21-46,213,215. Retrieved from https://www.proquest.com/scholarly-journals/indian-labour-migration-arab-gulf-states-impact/docview/906101363/se-2?accountid=8330

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika, 68*(3), 589-599. doi:10.1093/biomet/68.3.589

Raymer, J., De Beer, J., & 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*(2), 185-215.

Raymer J., Guan, Q., & Ha, J.T. (2019). Overcoming data limitations to obtain migration flows for ASEAN countries. *Asia and Pacific Migration Journal, 28*(4): 385-414. DOI: 10.1177/0117196819892344

Raymer J., Guan, Q., Shen, T., Wiśniowski, A., & Pietsch, J. (2020). *Estimating International Migration Flows for the Asia-Pacific Region: Application of a Generation-Distribution Model*. Work-in-progress project paper. School of Demography, Australian National University: Canberra, Australia.

Raymer J., Liu N., & Bai X. (2019). Age Articulation of Australia's International Migration Flows. In Franklin R. (Eds), *Population, Place, and Spatial Interaction* (pp. 171-200). Singapore: Springer. https://doi.org/10.1007/978-981-13-9231-3_10

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

Rogers, A. and Castro, L. J. (1981). *Model migration schedules.* Laxenburg, Austria: International Institute for Applied Systems Analysis.

United Nations, Department of Economic and Social Affairs, Population Division (2015). *Trends in International Migrant Stock: The 2015 Revision.* (United Nations database, POP/DB/MIG/Stock/Rev.2015).

United Nations, Department of Economic and Social Affairs, Population Division (2019). *International Migrant Stock 2019.* (United Nations database, POP/DB/MIG/Stock/Rev.2019).

United Nations, Department of Economic and Social Affairs, Population Division (2020). *International Migrant Stock 2020.* (United Nations database, POP/DB/MIG/Stock/Rev.2020).

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.

Wiśniowski, A., Forster, J. J., Smith, P. W., Bijak, J., & Raymer, J. (2016). Integrated modelling of age and sex patterns of European migration. *Journal of the Royal Statistical Society. Series A, (Statistics in Society), 179*(4), 1007–1024. https://doi.org/10.1111/rssa.12177

Appendix 1: Technical notes

This paper adopts Raymer et al.'s (2011) method to disaggregate origin-destination flows by age and sex. Their method starts from a complete OD (Origin by Destination) two-way table of migration flow with two main effects, (O_i) & (D_j), and one interaction, (OD_{ij}), $n_{ij} = (T)(O_i)(D_j)(OD_{ij})$, where n_{ij} is the observed flow of migration from origin *i* to destination *j* and *(T)* is the overall migration level. Then they incorporate age and sex information by multiplicative component model into a saturated ODAS four-way (origin by destination by age by sex),

$$n_{ijxy} = (T)(O_i)(D_j)(A_x)(S_y)(OD_{ij})(OA_{ix})(OS_{iy})(DA_{jx})(DS_{jy})(AS_{xy})(OAS_{ixy})(DAS_{jxy})(ODA_{ijx})(ODS_{ijy})(ODAS_{ijxy}).$$

Similar to Raymer et al.'s (2011) approach, we assume that recently produced origin-destination flow estimates from Raymer et al. (2020) are true and then add in age and sex main effect and two-way or higher dimension interaction. European countries have considerably more data regarding the age and sex pattern of emigration and immigration. In Asia-Pacific however, migration data are extremely limited and hardly include any age- and sex-specific details. Therefore, the aim of this paper is to extend previous work by including age and sex characteristics into country-to-country migration flows estimates for 53 Asia-Pacific populations.

There are four stages of the model, and each stage generates a set of stand-alone estimates that can be assessed. In Stage 1, a naïve age and sex schedule is assumed and combined with the OD table, $n^1_{ijxy} = (T)(O_i)(D_j)(A_x)(S_y)(OD_{ij})(AS_{xy})$. In the naïve model, we use the same age schedule for all countries in five-year age group, where the last age group is 85 years old and over. The overall sex effect (S_y) is set to one in the naïve assumption. The age component of the naïve schedule is generated from the official data of Australia, New Zealand and South Korea. To include uncertainty into the model, we adopt random subsampling technic (Efron 1981). Only 80 percent of the full data is used in each iteration and 1,000 iterations are performed, consistent with the previous OD estimates.

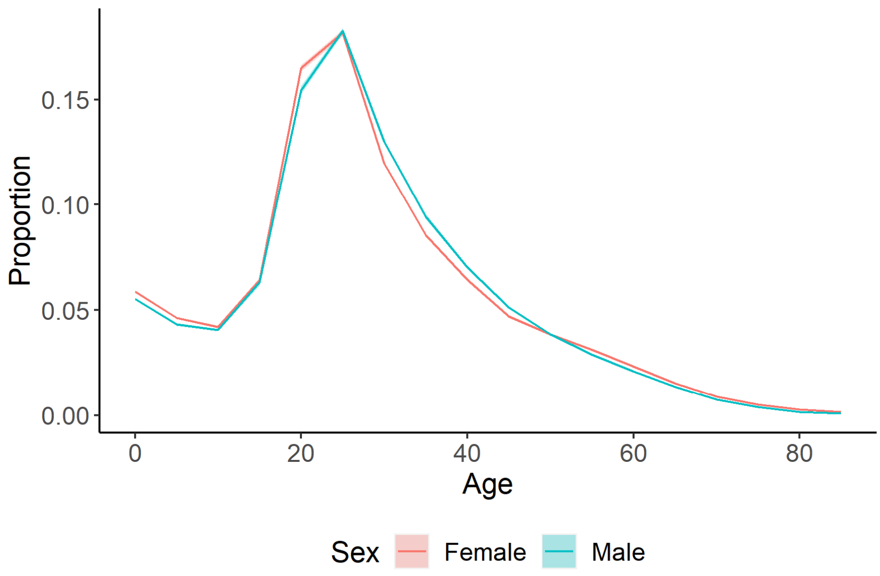


Figure A1. Naïve AS schedule

We calculated the age distribution of migrants by sex, year, and migration direction, and then use loess regression to smooth the distributions by sex and direction of migration. The sums of fitted proportions for both sexes are rescaled to one implementing the assumption of sex ratio of all migrants being one. The rescaled schedules are shown in Figure A1 with 80 percent predictive intervals (from 10 percentile to 90 percentile). It is evident that the results are not sensitive to the change in input data since 1,000 iterations do not produce wide uncertainty at this stage. In addition, this naïve age schedule solely relies on major receiving countries, but a nearly identical age and sex schedule of migration flows can be derived from based on IMEM estimates among 30 European countries, where there are some major sending countries such as Poland, Bulgaria and Romania. We then combine the naïve schedule in Stage

1 with the original OD table. Each iteration of age and sex schedule would be matched to one iteration in the OD table. Using an Iterative Proportional Fitting procedure (IPF), we provide all migration corridors with the same age and sex structure.

In Stage 2, we incorporate destination-specific information into the model. Migrants arriving in a country can be native-born (nationals) or foreign-born (foreigners) with the age and sex distribution for these two groups likely being different. For example, about 40 percent of foreigners arrived in Korea between 2000 and 2018 were female. On the contrary, more than 50 percent of nationals returned to Korea were female. Therefore, we estimate separate destination by age by sex (DAS) tables for the native-born (nationals) and foreign-born (foreigner) populations and combine them using weights to obtain the full DAS table for each destination.

We first calculate the DAS table for foreign-born (foreigners) from the UN migrant stock. The UN migrant stock data provide stock figures by birthplace predominantly from 1990 to 2020 by age and sex for 52 destinations in Asia-Pacific excluding Taiwan (see Figure 1). As mentioned in the Data section, the age and sex structure of the stock may not reflect the one for flow, so we need to transform than into change in stock to simulate the flow statistic. The change in stock need to use a reverse cohort component projection to back project the stock five years younger in five years ago. Then the difference between the back projection value and the actual stock is the change in stock.

Table A1 shows the five-year aggregated UN stock by period and age in Australia for both sexes. To calculate the change in stock, we examine this table in cohort perspective, namely diagonally. For example in Table A1 panel a, there are 109,890 foreign-born migrants at age 5-9 in Australia from the UN data in 2020. Assuming that no migrant moved in or out of the population after age 0, the 109,890 foreign-born migrants should be the survivor of 110,319 foreign-born migrants in 2015 at age 0-4, which is calculated from the survival function of UN life table (Table A1 panel b) in 2015 at age 0-4 ($109{,}890/0.9961 = 110{,}319$). Using panel a migrant stock divided by the cell five year and five age earlier in panel b yields the result of expected starting migrant in panel c. However, only 65,325 migrants at age 0-4 in 2015. Shown in panel d, subtracting panel c to panel a directly gives the change in stock 44,994 ($110{,}319 - 65{,}325$). This change in stock of is the minimum number of migrants enter Australia between 2015 and 2020 and they are first-censored in 2020 at age 5-9. First-censored refers to the people not observed in the 2015. This change in stock is a minimum flow since it is based on the assumption that no migrant exit the population. We add row age 0 in panel d to reflect the definition of the change in stock.

Figure A2 illustrates the mechanism and limitations of this minimum flow approach. There are six typical cases of the migrants' entry timing and age as well as their lifetime spent in the population between two consecutive periods. The first three persons are all censored in 2020 at age 5-9, but only the second and the third person fit into the first-censored requirement, as in the first person was counted in 2015 at age 0-4. Although the second and the third ones are both 5-9 in 2020, they could enter the population at either age 0-4 or 5-9. Thus, we assume that the probability of entering into two consecutive age groups are the same. As a results, half of the change in stock is expected to enter at age 0-4 and the other half enter 5-9. However, the assumption of non-leaving does not hold in reality and miss some newcomers. For example, the fifth line can be viewed as one continuous line with the second or the third line. Their ages differ in Figure A2, but in term of age group, one person is observed at age 0-4 in 2015, while another one is observed at age 5-9 in 2020. The stock data cannot differentiate these two persons so they would be treated as the same person, and consequently not included in the minimum change for not meeting the first-censored requirement. In this simplified example, three persons were observed in 2020 at age 5-9, but only one of them is regarded as a flow in the change in stock.

Table A1. Australian data from UN (both sexes)

a. UN migrant stock

	1990	1995	2000	2005	2010	2015	2020
0	33,998	23,910	29,810	35,420	57,590	65,325	
5	92,271	79,720	74,540	91,330	129,920	149,455	109,890
10	129,260	126,300	119,220	128,200	160,830	188,687	197,666
15	163,141	165,170	171,150	191,430	224,630	255,477	427,067
20	236,668	228,200	230,900	290,020	443,540	433,434	663,110
25	328,204	299,060	282,550	320,230	520,520	589,939	742,345
30	380,707	394,420	351,000	361,350	469,690	645,786	636,572
35	385,653	409,620	433,230	414,060	462,140	562,320	641,552
40	427,063	412,710	432,920	477,430	486,860	528,103	583,497
45	380,996	452,140	448,670	429,560	475,750	523,600	539,349
50	311,763	346,810	421,580	458,300	517,120	533,248	531,157
55	262,892	300,060	335,690	447,110	431,310	482,102	546,457
60	233,364	253,900	294,040	335,290	440,210	433,444	487,216
65	215,590	224,510	240,520	286,370	329,210	434,065	423,859
70		192,170	205,430	226,260	270,340	319,469	406,857

b. Survival function (Sx) from UN life table

	1990	1995	2000	2005	2010	2015	2020
0	0.9916	0.9929	0.9937	0.9945	0.9954	0.9961	
5	0.9992	0.9993	0.9994	0.9995	0.9996	0.9996	
10	0.9982	0.9983	0.9987	0.9989	0.9990	0.9991	
15	0.9964	0.9964	0.9972	0.9977	0.9981	0.9982	
20	0.9958	0.9957	0.9966	0.9973	0.9977	0.9979	
25	0.9956	0.9954	0.9962	0.9968	0.9972	0.9974	
30	0.9948	0.9948	0.9955	0.9960	0.9963	0.9966	
35	0.9934	0.9936	0.9942	0.9946	0.9949	0.9953	
40	0.9904	0.9912	0.9916	0.9921	0.9926	0.9932	
45	0.9845	0.9865	0.9878	0.9882	0.9889	0.9899	
50	0.9744	0.9781	0.9811	0.9825	0.9833	0.9849	
55	0.9576	0.9640	0.9694	0.9730	0.9748	0.9772	
60	0.9307	0.9409	0.9505	0.9569	0.9609	0.9648	
65	0.8899	0.9035	0.9190	0.9303	0.9370	0.9432	

c. Expected starting migrant by back projection

	1990	1995	2000	2005	2010	2015	2020
0	80,395	75,077	91,908	130,639	150,139	110,319	
5	126,403	119,302	128,272	160,908	188,770	197,744	
10	165,461	171,446	191,682	224,877	255,722	427,436	
15	229,019	231,743	290,830	444,552	434,273	664,278	
20	300,309	283,761	321,314	521,950	591,302	743,910	
25	396,163	352,620	362,721	471,199	647,587	638,207	
30	411,746	435,494	415,917	464,015	564,407	643,748	
35	415,448	435,721	480,223	489,517	530,815	586,248	
40	456,511	452,675	433,182	479,540	527,520	543,034	
45	352,255	427,347	463,969	523,300	539,256	536,592	
50	307,950	343,204	455,712	438,981	490,274	554,849	
55	265,147	305,008	345,878	452,428	444,656	498,575	
60	241,240	255,617	301,292	344,051	451,728	439,317	
65	215,943	227,368	246,192	290,597	340,932	431,344	

d. Change in stock

	1990-1995	1995-2000	2000-2005	2005-2010	2010-2015	2015-2020	
0							
5	46,397	51,167	62,098	95,219	92,549	44,994	
10	34,132	39,582	53,732	69,578	58,850	48,289	
15	36,201	45,146	72,462	96,677	94,892	238,749	
20	65,878	66,573	119,680	253,122	209,643	408,801	
25	63,641	55,561	90,414	231,930	147,762	310,476	
30	67,959	53,560	80,171	150,969	127,067	48,268	
35	31,039	41,074	64,917	102,665	94,717	-2,038	
40	29,795	26,101	46,993	75,457	68,675	23,928	
45	29,448	39,965	262	2,110	40,660	14,931	
50	-28,741	-24,793	15,299	93,740	63,506	12,992	
55	-3,813	-3,606	34,132	-19,319	-26,846	21,601	
60	2,255	4,948	10,188	5,318	13,346	16,473	
65	7,876	1,717	7,252	8,761	11,518	5,873	
70	353	2,858	5,672	4,227	11,722	-2,721	

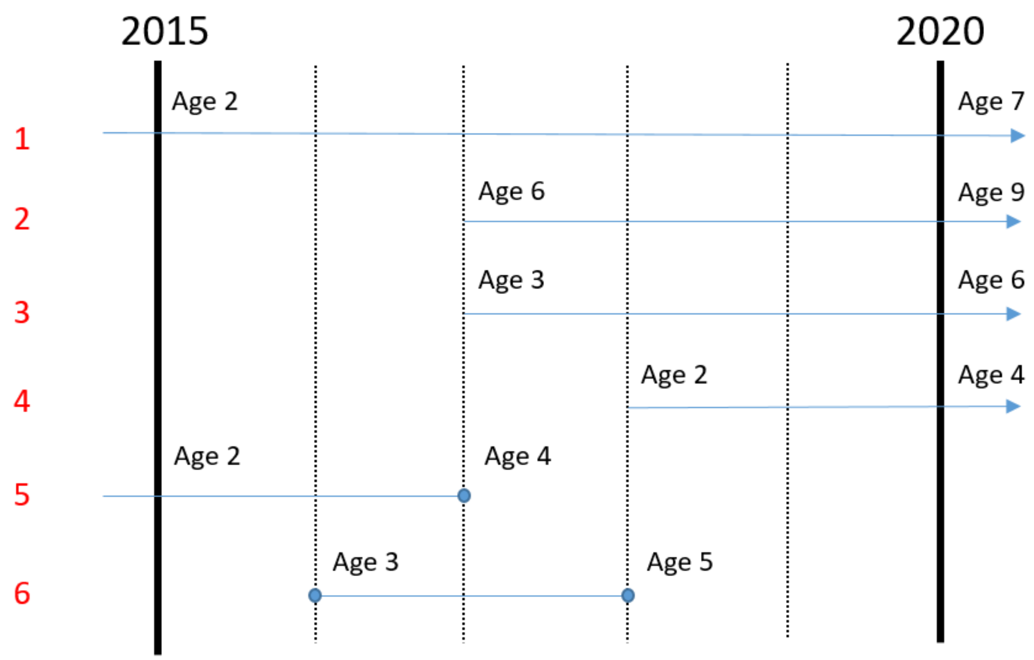


Figure A2. Census timing and possible migrants

Table A2 presents the results from the foregoing method. For the migrants entering Australia between 2015 and 2020 and first-censored at age 5-9, we assume that half of them ($\frac{44{,}994}{2} = 22{,}497$) enter in beginning of this period (2015) at 0-4 and the other half right at the end of the period (2020) at age 5-9. Except the first and last age groups and years, other cells are comprised by two components (shown in Table A2). For example, the sum of a half from 2015-2020 at age 10-14 ($\frac{48{,}289}{2} = 24{,}145$) and the other half from 2010-2015 at age 5-9 ($\frac{92{,}549}{2} = 46{,}274$) is the 70,419 ($24{,}145 + 46{,}274$) flows in 2015 at age 5-9 in table 2. However, in the model, instead of a deterministic half of the migrants entering at beginning of the period and half at the end of the period, we provide a normal distribution of mean at 0.5 and standard deviation at 0.2. Therefore, when it is 0.5 the value would be as provided in the example. With this set of mean and standard deviation, 95 percent of iterations would yield a weight between 0.1 and 0.9, which means between 10% and 90% of migrants entering at the beginning of the period and the corresponding rest at the end of the period.

Table A2. Result of the cohort method for Australia

	1990	1995	2000	2005	2010	2015	2020	Median
0	23,199	25,584	31,049	47,610	46,275	22,497		28,316
5	17,066	42,990	52,450	65,838	77,035	70,419	22,497	52,450
10	18,101	39,639	56,022	75,205	82,235	148,800	24,145	56,022
15	32,939	51,387	82,413	162,792	153,160	251,847	119,375	119,375
20	31,821	60,720	78,494	175,805	200,442	260,060	204,401	175,805
25	33,980	58,601	67,866	120,692	179,499	98,015	155,238	98,015
30	15,520	54,517	59,239	91,418	122,843	62,515	24,134	59,239
35	14,898	28,570	44,034	70,187	85,670	59,323	-1,019	44,034
40	14,724	34,880	13,182	24,552	58,059	41,803	11,964	24,552
45	-14,371	2,328	27,632	47,001	32,808	26,826	7,466	26,826
50	-1,907	-16,174	4,670	-2,010	33,447	42,554	6,496	4,670
55	1,128	568	3,291	19,725	-2,987	-5,187	10,801	1,128
60	3,938	1,986	6,100	9,475	8,418	9,610	8,237	8,237
65	177	5,367	3,695	5,740	10,242	4,399	2,937	4,399
70		177	1,429	2,836	2,114	5,861	-1,361	1,771

We did not include migrant stock age 0-4 though they must enter Australia at age 0-4. The reason is that the raw stock number is incomparable with the change in stock we calculate in cohort method. Because 0-4 is the initial census time, all people recoded at this age group meet the first-censored requirement. As a result, this stock number are likely to be much larger than other changes in stock.

Then we calculate the median for each row to obtain the age structure. Using median rather than mean is to avoid extreme values in certain cohorts. There are a few negative change in stock, so using median could minimize their effects on the results. In the result of Australia, there is no negative values, but for other countries, the small population in particular, could yield negative flows (shown in Figure A3). Abel and Cohen (2019) mentioned two ways to deal with the negative values when they use stock differencing to estimate flow: 1) treating negative as zero; 2) reverse the negative to positive. However, we are using the age distribution rather than these median number directly and the shapes of most countries are more or less matching the naïve schedule in Figure A1 with a labour force peak, so we merely shift up the entire age schedule and set the lowest value of each country to zero.

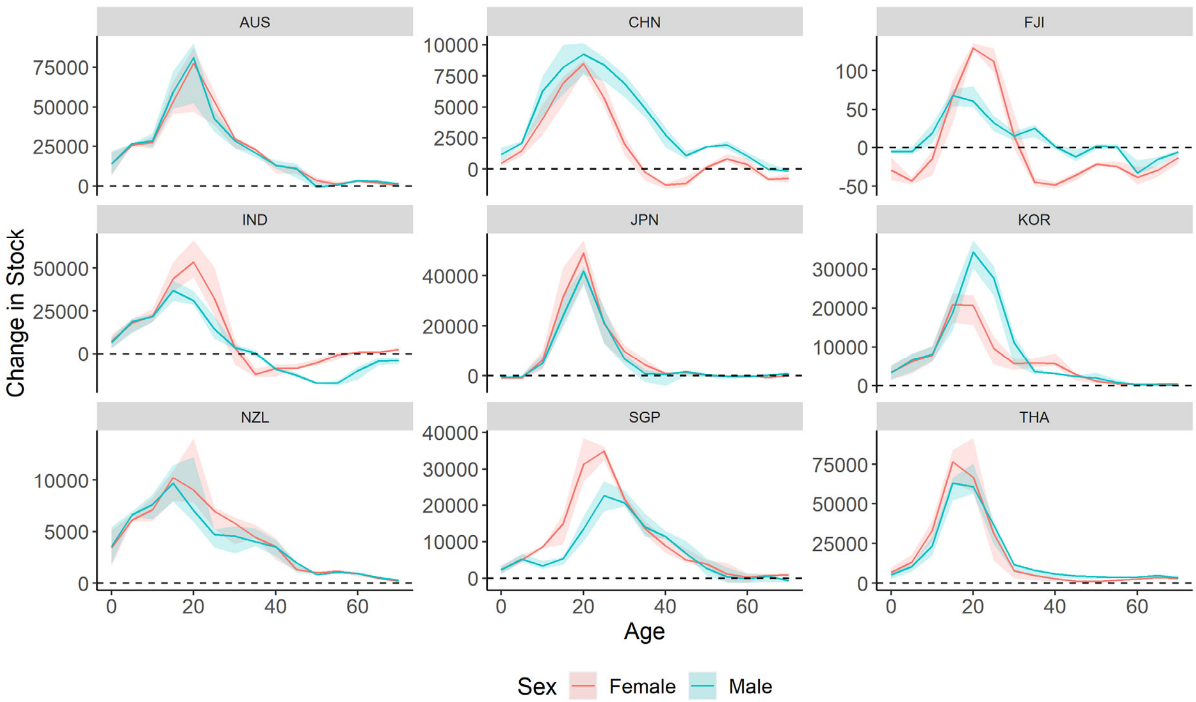


Figure A3. Change in stock by sex (before adjustment)

Another irregularity of this result is some uptick or fluctuations at older ages. Additional, the stock data can only provide information until age 70. We present the original results in dashed line in Figure A4. We rely on empirical regularities of migration suggested in Rogers and Castro (1981). Our first attempt to smooth the curves and extend older age groups (75-79, 80-84, and 85+) was by fitting a nonlinear least squares model, which was proposed by Rogers and Castro (1981), with pre-labour force component and labour force component. The formula for the migration proportion at age x is:

$$p(x) = a_0 + a_1 \exp(-\alpha_1 x) + a_2 \exp\{-\alpha_2(x - \mu_2) - \exp[-\lambda_2(x - \mu_2)]\},$$

where first term in the addition is a constant, and the second one is the pre-labour force decent, and the last one is the peak of labour force.

However, nonlinear modelling requires initial values for parameters to start the estimates and only good initial values could lead to optimal solutions. As a result, we simplify this formula and only keep the exponential distribution for age groups over μ:

$$y = a \exp(-b(x - \mu)), \qquad \text{(Eq. 1)}$$

where y is the number of number of change in stock and x is age.

Three parameters, a, b and μ, are needed to generate smoothed y. Since only the post-labour force descent would be adjusted, we simply set μ at 25, 30 or 35. As for a, it follows the age of μ and the exponential curve starts from value at age μ. For example, when μ is 25, a is the change in stock at 25. For b, the rate of descent, we calculate the rate of change from age 25 to 70 with the gap of 15 years of age between each group. For example, one of the rates of change would be comparing age 25 to age 40, and the other is 30 to 45. The rationale behind the 15-year of gap is to balance the unwanted fluctuations and the number of rates available for the average. If the gap is shortened to 10, the rate may not reflect the overall trend. On the other hand, if the gap is widened to 20, we only have six rates of change. Then all the rates of change are averaged by geometric mean excluding those negative rates. Using Eq. 1, we can fit values older than μ, and obtain three sets of estimates. Using sum of square error over age 40 between these estimates and the original data, we determine which μ and a is best capturing the data. Therefore, the smooth curve could start out from 25, 30 or 35 (where solid lines and dashed lines diverge in Figure 3). Noted that, the shapes of the change in stock for female migrants in Iran is unable to smooth, we hence supply the naïve curve to them.

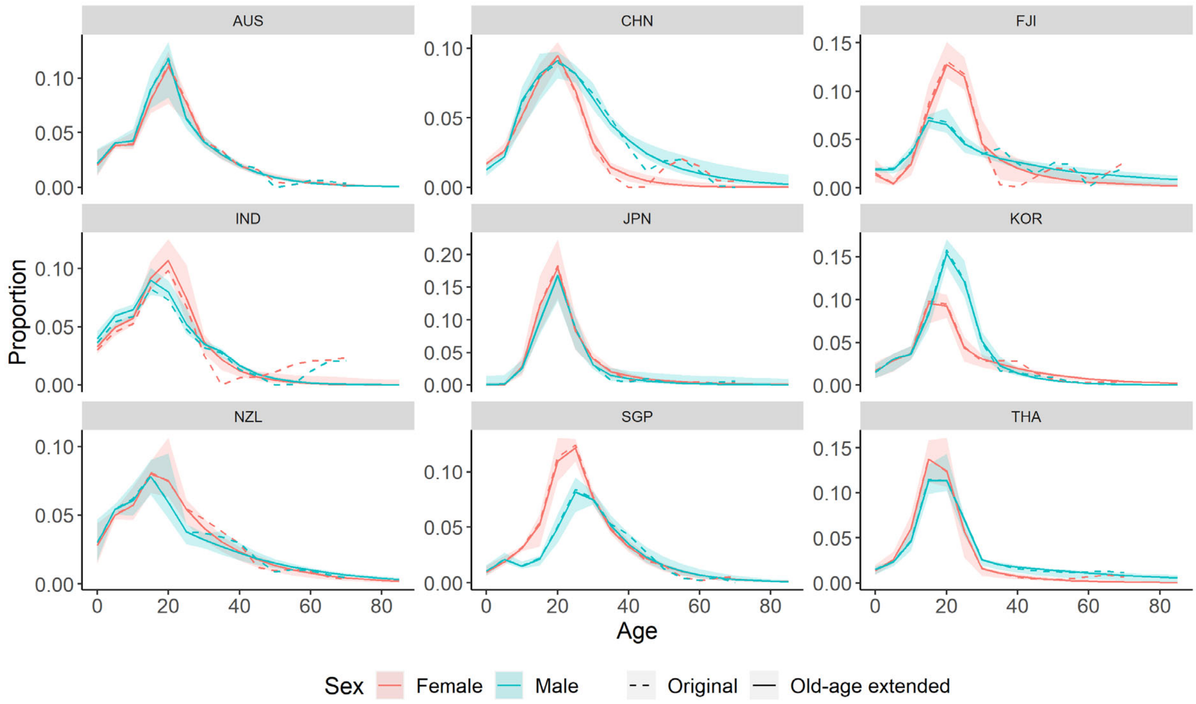


Figure A4. Change in stock by sex (adjusted)

After obtaining the age and sex distribution of foreign-born (or foreigner) arrival, we construct the one for the native-born (or national) arrival borrowing reported data from Australia and Korea. These two countries are only ones provide detail statistics on migration of the native-born (Australia) and nationals (Korea) (hereinafter referred to as "nationals"). The national naïve AS schedule is calculated using the same method in Stage 1 --- loess regression to smooth the age distributions of arriving nationals by sex. To estimate the sex ratio for nationals, the foreigner flows from OECD (2021) is used. OECD has outflow of foreigner by sex and destination from Australia, Korea and New Zealand. Assuming that the sex ratio of the outflow of foreigner is equal to the inflow of national of the destination, we derive the sex ratio of nationals arriving. The magnitude of flows from Australia, Korea and New Zealand differs, so we rescale them to the same weight and calculate the average sex ratio of return flow to each country. Similarly, only 80% of the data is randomly selected when estimating naïve AS schedule and sex ratio national arrivals. The results are shown in Figure A5.

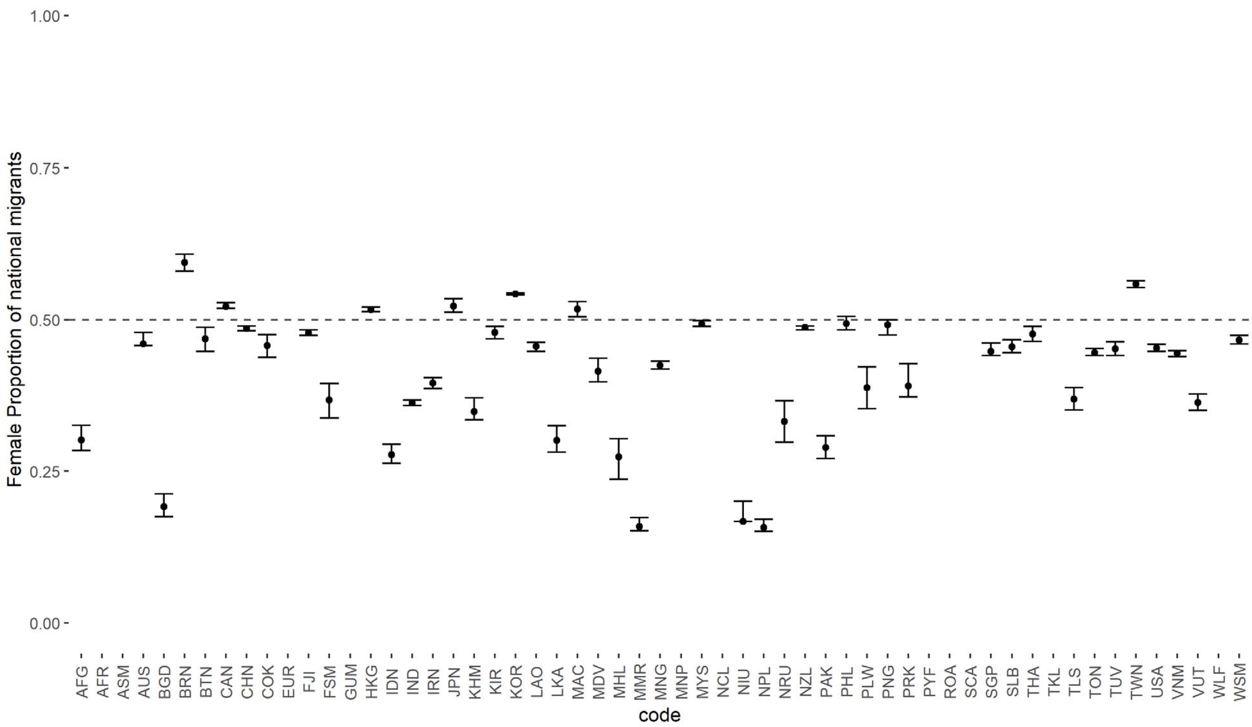


Figure A5. Female proportion of national migrant returning

With the proportion of foreigner and national arrivals by age and sex, the last step for Stage 2 is to combine the foreigner and national DAS into one by weights. To do this, we estimate the proportion of migrants who are nationals or native-born from known sources. Australia, South Korea and New Zealand statistical office provides the proportion of migrants that are nationals. Australian data adopt the country of birth definition, South Korea citizenship definition and New Zealand prior residency. Since we do not have the luxury to distinguish these definitions, we treat them as the same. Again, 20% of data was removed randomly for each iteration. From the proportion of national over total migrants in these three countries, we need to predict the proportions for the others. We assume logistic function to estimate their proportions,

$$\text{pro}_j = \frac{1}{1+e^{k_j x_j}}. \qquad \text{(Eq. 2)}$$

Since there is no empirical relations that we can draw on, net migration rate, effectiveness of migration and logged proportion of migrant stock (to transform the domain to all real numbers). To some extent, these three variables (j) are related to the exchange of migrants and the impact of migration. For each variable, we obtain the k based on the geometric mean of the three countries. For instance, plunging in the observed proportion and net migration rate of Australia to Eq. 2, we have a k for net migration rate. After we do this for three countries, we get three ks for net migration rate. Since k is exponential in natural, we average them by geometric mean and obtain k_j (in this case j is net migration rate). With this k_j and net migration rate from the 53 countries, we can calculate the proportion for each countries. The same procedure is applied to other two js, namely effectiveness and logged proportion of migrant stock. At the end, each country has three pro_j. We calculate the arithmetic mean of the three to obtain final estimates on the proportion of nationals in total migrants (refer to Figure A6).

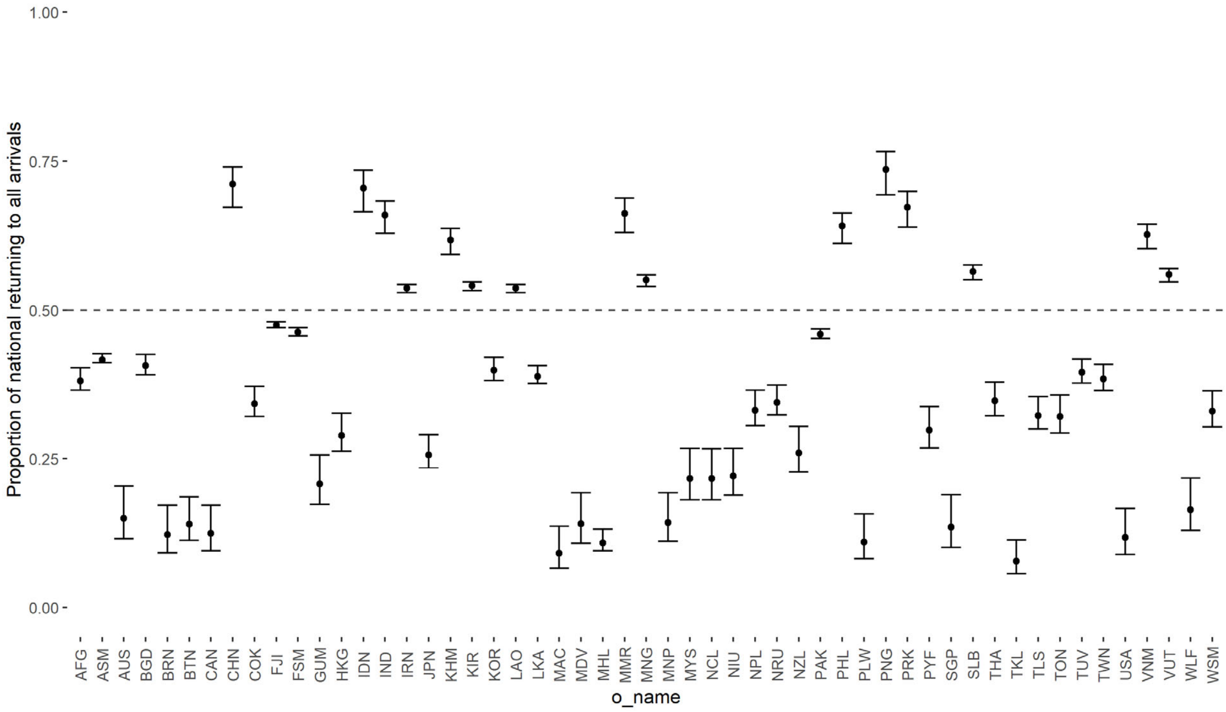


Figure A6. Proportion of national returning to all arrivals

After the foregoing steps, we can add up these two DAS tables by the weights of each countries. Our new model now include three more interaction terms, where three two-way interactions, (AS^*_{xy}), (DA_{jx}) and (DS_{jy}) automatically derived from the DAS table, $n^2_{ijxy} = (T)(O_i)(D_j)(A_x)(S^*_y)(OD_{ij})(AS^*_{xy})(DA_{jx})(DS_{jy})(DAS_{jxy})$. One issue with this log-linear model is that the sex component from (DAS_{jxy}) differs from the previous sex main effect (S_y). Therefore, we have to decide which overall sex ratio we would use in the result. Since the previous sex ratio at unity is purely naïve, we opt for the sex ratio (male to female) from DAS table and derive (S^*_y), which is between 1.10 and 1.16 with a median at 1.13. On the other hand, the age main effect (A_x) is kept as it is based on the reported data. The justification for keeping the naïve (A_x) is to adjust for the pre-labour force peak since the change in stock method underestimates age 0-4. (AS^*_{xy}) is also changed as a result of the new (S^*_y).

Stage 3 is to add in OS interactions (OS_{iy}) to all 57 populations and DS interactions (DS_{jy}) for the four Rest of the World regions from auxiliary information. In IMEM estimates of migration flows among European countries (Wiśniowski et al., 2016), the correlation of age profile between emigrants and immigrants reaches 0.97. Since the correlation of age profile between emigrants and immigrants in the Stage 2 estimates is already 0.90, we choose not to adjust the (OA_{ix}). On the other hand, the correlation of proportions of female between emigrants and immigrants in IMEM estimates is only 0.15. Low correlation proportions of female is also found in Abel's (2018) estimates, at about -0.09. Therefore, we borrow the OS and four DS interactions from Abel's estimates (2018) to enhance our results. In his study, migration flows by origin, destination and sex are estimated for most countries in the Asia-Pacific. However, Abel's estimates include many zeros and 30 percent of none-zero flows are highly skewed regarding sex (single sex accounting for more than 80% of the flow). Due to these limitations, we only borrow Abel's OS interaction rather than the detailed three-way interaction (ODS).

Figure A7 presents the proportion of female immigrants by destination with 80 percent predictive intervals (from 10 percentile to 90 percentile). The black dots are from Stage 2 estimates and triangles are from Abel's (2018) estimates. Two sets of estimates are not particularly close, but ours have fewer extreme values. Abel's estimates also misses out a number of Pacific countries. However, the sex ratios of four rest of the world (ROW) regions, namely Africa, the rest of Asia, Europe and South and Central America, were set at one in Stage 2. This assumption overlooks the effect of Middle East attracting substantial number male manual labour. Therefore, we combine Abel's DS and DS from Stage 2 by a proportion to generate a new proportion for these four ROWs, namely AFR, EUR, ROA and SCA. The proportion is assumed a normal distribution with the mean at 0.5 and standard deviation at 0.2. The reason we choose not to apply Abel's DS directly was that their estimates subject to more extreme cases, so we take a more conservative approach. The red error bars are the new proportion of female immigrants by destination after adjusting the ROWs sex ratio. Since the DS of other countries is fixed, the sex main effect (S_y^{**}) only increases slightly to a median at 1.17 and most countries cannot see Stage 2 estimates because they overlap with Stage 3.

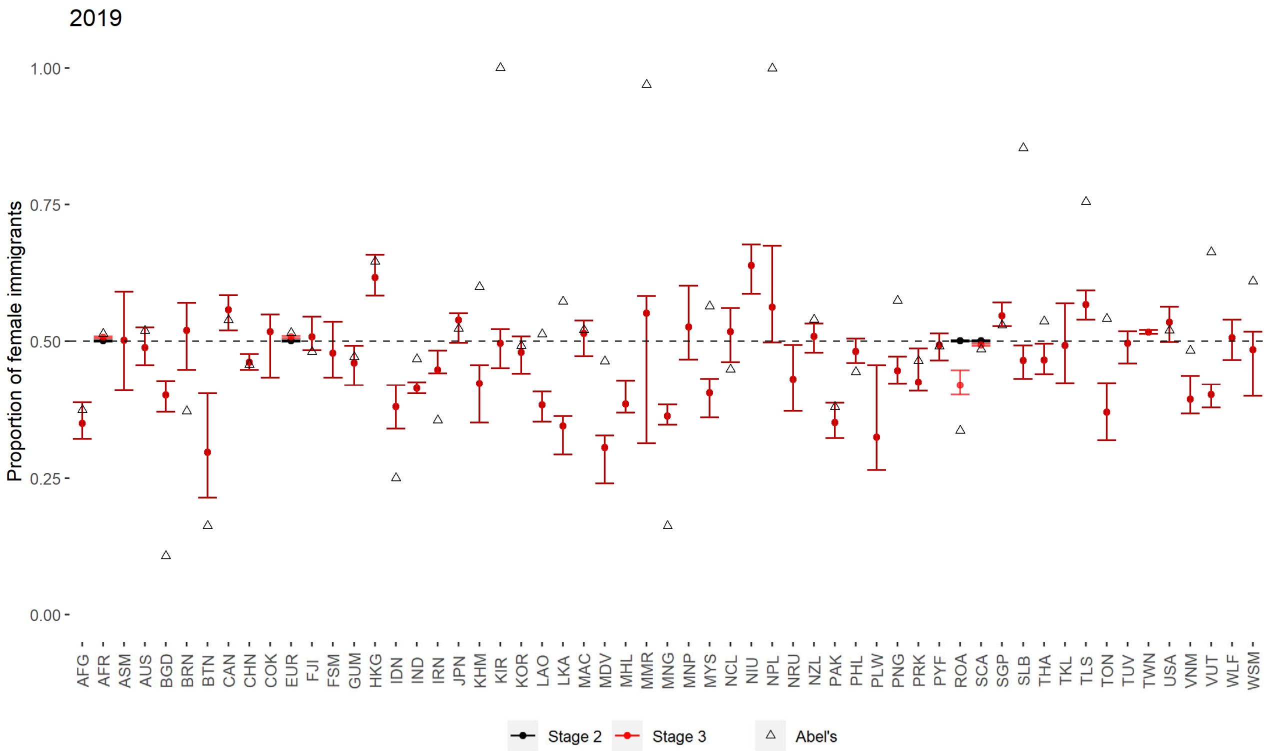


Figure 5. Proportion of female immigrants by destination and sex

Figure 2 shows the proportion of female emigrants by origin with 80 percent predictive intervals (from 10 percentile to 90 percentile). The error bars are from Stage 2 estimates and triangles are from Guy Abel (2018) estimates. Since results from Stage 2 does not include assumptions in OS interaction, and black error bars are mere artefact of the interactions of DAS and OD. We hence combine Abel's OS with OS from Stage 2 estimates using the same normal distribution as weight as above to avoid extreme values (see Table 1). Controlling for the new sex main effect (S_y^{**}), the OS becomes the red error bars.

After obtaining (DS_{jy}^{*}) and (OS_{iy}), we run log-linear model, $n_{ijxy}^{3} = (T)(O_i)(D_j)(A_x)(S_y^{**})(OD_{ij})(AS_{xy}^{**})(DA_{jx})(DS_{jy}^{*})(OS_{iy})(DAS_{ijy}^{*})$ to combine all effects. Noted that (DAS_{ijy}^{*}) is different as well due to the modifications in (DS_{jy}^{*}).

We compare the result from Stage 3 with the relatively good data in Figure 3 and 4. The reported value is exactly for that year and Stage 3 estimates are value of that year with 80 percent predictive intervals, however Abel's estimates is the average of three point estimates in 2000, 2005 and 2010. Australia, Korea and New Zealand have detail age and sex components for both emigrant and immigrant, while Canada only has the number of immigrant and emigrant by sex. Data are obtained from national statistics offices of these four countries (ABS 2020; Statistics New Zealand 2020; Statistics Korea 2020; Statistics Canada 2020b). "Although these data are reliable, none of their definitions is perfectly align with the 1998 United Nations recommendations. Australia defines immigrants and emigrants as persons who stay in or leave Australia for 12 months or more over a 16-month period, regardless of citizenship or birthplace (ABS 2020). New Zealand defines immigrants and emigrants as permanent and long-term arrivals or departures who arrived for a continuous stay of 12 months or more or departed for 12 months continuously or more, regardless of citizenship or birthplace (Statistics New Zealand 2020). South Korea defines immigrants and emigrants as persons who have continuously resided in or left the country for more than one year, regardless of citizenship or birthplace (Statistics Korea 2020). Canadian immigration and emigration statistics are recorded separately for permanent (including citizens) and temporary residents (Statistics Canada 2020a & 2020b), with no clear measurement on duration of residence and only net migration numbers are provided for temporary migrants." (Raymers et al. 2020)

Since Stage 3 incorporates Abel's estimates of migrant flows by sex, Figure 3 also includes Abel's results in squares. Figure 3 compares the sex component of Australia, Canada, Korea and New Zealand. For the immigrants (DS), the estimated proportions of Australia and New Zealand in Stage 3 and Abel's study are in different side of the reported. However, Stage 3 estimates is slightly more conservative bringing it closer to the reported data. Since both estimates are derived from the stock, the dominant sex of the migrants is the same even though the ratios differ. Regarding the emigrants (OS), the Stage 3 is closer to the reported, but still Canada results flipped to the other side. Overall, the Stage 3 Estimates appears to perform better than Abel's estimates. For Figure 4, it shows the age and sex component for immigrants to and emigrant from Australia, Korea and New Zealand. Although the estimates do not follow the reported data perfectly, the peaks in particular, the overall trends however are mostly capture.

Therefore, in Stage 4, we replace the estimates with reported data for the four countries with relatively good international migration statistics by age and sex. Using IPF and maintaining the main sex and age effect, other countries should be adjusted accordingly. However, since the change is minor compared to the overall level, the effect on other countries is very limited.