Estimating Migrant Population in Malaysia

Emir Izat bin Abdul Rashid

Asia School of Business

Outline

- Introduction and Motivation
- Preview
- Model
- Population Prediction
- Conclusion

Research Preview

- How can we improve estimation of migrant population in Malaysia?
- Issue: high levels of undocumented population can undermine social welfare e.g., labor market, public health, etc.
- During COVID-19 period, Malaysian government launched a vaccination drive and committed to not take legal action against the undocumented.
- Can use this event to estimate the population of migrant population.
- Developed a model that attempts to endogenize migrant's choice in vaccinating.
- Using different parameter spaces and assumptions, we can predict the migrant population as well as highlighting the mechanism that drives vaccine turnout.

Why Need to Estimate Migrant Population?

- We have limited information in the effect of migration to a range of outcomes in Malaysia.
- For example, public health policies are less effective at addressing epidemics. Housing supply may not catch up to migrant inflow. Labor market effects.
- Collecting information on macro-level migrant population is important to deliver effective policies.
- Micro-level population may be useful but must account for internal movement which happens a lot in migrant communities.
- Cannot proceed further with observational work without reliable data.

Literature

- Theoretical implications in empirical methodology. Izzo et. al (2022); Slough and Tyson (2024); Bueno de Mesquita and Tyson (2024); Ashworth et. al (2017)
- Economics of migration. Mobarak et. al (2023); Boustan et. al (2020); Tabellini (2020)

Overview of Migration in Malaysia

- Government regulates certain nationalities to work in certain occupations. No information on compliance but can assume that these restrictions are not tightly enforced.
- Large undocumented population: limited information on estimates. 2 to 4 million according to UNHCR, how true is this estimate?
- Official statistics: 182,820 undocumented.
- Migrant population centers around two regions: Sabah and the Klang Valley.

Example of Migrant Worker Community

Figure 1: Example of a *Kongsi*

Estimation Strategy

- Malaysian government committed to not take legal action against migrants who vaccinate. Why? Public health concerns outweighs potential national security concerns.
- Data for daily vaccination drives are publicly available. How to use this data to estimate the migrant population?
- Simplified model can clarify the mechanisms that drive vaccine turnout among migrant. Using this model, we can find the probability of vaccination for certain migrant groups thus estimating the total population.

Model Description

- Migrant i of legal status $\tau = \{l, u\}$ and vaccine preference $\nu = \{v, av\}$ have to choose whether to vaccinate. Legal status is either legal l or undocumented migrant u. Vaccine preference is whether they are comfortable with receiving vaccine.
- If $\tau = u$, vaccination carries a risk of getting caught with $Pr(\text{caught}) = p$ i.e. we take this as exogenous. Getting caught will incur a cost c.
- If $\nu = av$, vaccination induce a reduction in utility r_{av} and if $\nu = v$, not vaccinating induce the utility by r_v.
- If migrant do not vaccinate, there will be an idiosyncratic shock $\eta_i \sim F$ where F is a CDF with support ranging from $(0, +\infty)$ that conveys the level of severity of COVID-19[1].

[1] One can interpret this from no symptoms to death

Utility Function of Migrant

- Utility function for a legal migrant is

$$U_l = \begin{cases} -1_{\nu=av} r_{av} & \text{If vaccinate} \\ -1_{\nu=v} r_v - \eta_i & \text{If not vaccinate} \end{cases}$$

- Utility function for an undocumented migrant is

$$U_u = \begin{cases} -1_{\nu=av} r_{av} - pc & \text{If vaccinate} \\ -1_{\nu=v} r_v - \eta_i & \text{If not vaccinate} \end{cases}$$

- Utility function is simple to illustrate the inherent selection bias in terms of those who choose to vaccinate.

General Outcomes of the Model

- If $\tau = l, \nu = v$, probability of vaccination is 1.
- If $\tau = u, \nu = v$, probability of vaccination is $Pr(\eta_i \geq pc - r_v)$
- If $\tau = l, \nu = av$, probability of vaccination is $Pr(\eta_i \geq r_{av})$
- If $\tau = u, \nu = av$, probability of vaccination is $Pr(\eta_i \geq pc + r_{av})$

How Many Will Vaccinate?

- Thus, suppose that $N_{\tau\nu}$ is the population for τ, ν:

$$N_{\text{migrant, vaccinate}} = N_{lv} + Pr(\eta_i \geq pc - r_v)N_{uv} +$$
$$Pr(\eta_i \geq r_{av})N_{lav} + Pr(\eta_i \geq pc + r_{av})N_{uav}$$

$$N_{\text{migrant, no vaccinate}} = 0 + F(pc - r_v)N_{uv} +$$
$$F(r_{av})N_{lav} + F(pc + r_{av})N_{uav}$$

Assumptions Needed For Identifying Groups of Migrants

- We need to assume the parameter values p, r, c. For example, if $p = 0$ as promised by the government, every vaccine-friendly migrant will vaccinate.
- We need to then assume the distribution of F in terms of COVID-19 severity. For example, since we expect long COVID to affect 10% of patients and lower when people vaccinate, we can employ gamma distribution with a left-skew.
- Last but not least, we need to assume the portion of migrants who are vaccine-preferred and vaccine-averse. For example, given parameter values, we can estimate the population of vaccine-averse nationals; if we assume the proportion for migrants is the same we can derive this.

Assumptions for Benchmark Estimate of Migrant Population

- Benchmark estimate is when $r_{av} \to \infty$ i.e. the case where $\nu = av$ for all migrant types do not vaccinate (anti-vaxxers).
- Also, $p = 0$ i.e. government promises that they will not catch undocumented migrants which means $\nu = v$ for all migrant types vaccinate for all $r_v \geq 0$.
- Lastly, assume that the vaccine preference is not correlated with identity i.e. I take many random sample of any sub-population, it will return roughly the same proportion of $\nu = v$ and $\nu = av$

How to Estimate This?

- Since we know that all three assumptions above are met, we can use Malaysian nationals to compute the expected proportion of ν for l, u, which will all equal each other.
- Then, using this expected proportion combined with the population all legal migrant, we derive the population of legal and undocumented migrant who are $\nu = v$.
- Finally, we derive the population of migrant who are $\nu = av$ which enables us to derive an estimate of population of undocumented migrant.

Result of Benchmark Estimate

During the vaccination drive, 24,096,130 Malaysians and 3,723,736 non-Malaysians received at least one-dose of COVID-19 vaccine.

Population	$\nu = v$	$\nu = av$	Total population
Malaysians	24,096,130	6,102,070	30,198,200
Documented Migrants	1,994,991	505,209	2,500,200
Undocumented Migrants	1,728,745	349,323	2,077,068
Total Migrants	3,723,736	854,533	4,578,269

Table 1: Migrant Estimates

Note: Red color indicates a value that is derived.

As one can observed, benchmark estimate is the lower bound of the UNHCR estimate.

Complications Arising From Parameter Changes: Vaccine Preference

- Suppose benchmark assumptions but we let $r_{av} = 0$, which means every migrant will get vaccinated.
- This implies total undocumented migrants will be 1.2 million people. The most conservative estimate of the undocumented population: difference is 800,000 people with the benchmark case.
- As r_{av} decreases, more $\nu = av$ will get vaccinated as the severity outweighs their preferences. This will cause our estimate of undocumented population to decrease.

Complications Arising From Parameter Changes: Non-Zero Probability of Getting Caught

- Assume now as the benchmark case but $p \neq 0, c, r_v$ such that $pc - r_v > 0$. We know that $\nu = av$ does not vaccinate. $Pr(\eta_i \geq pc - r_v >) > 0$ which means estimate will be higher than benchmark. Conditions to get the UNHCR upper bound?
- To obtain UNHCR upper bound, we first assume that population of undocumented is 4 million, of which 3.2 million are $\nu = v$.[2]
- Then, 1.728 million is observed in data which implies that $Pr(\eta_i \geq pc - r_v) = 0.54$ i.e. given the undocumented migrant prefers vaccine 46% of them are deterred from receiving them due to fear of being caught.

[2] From the proportion assumption based on Malaysian nationals

Complications Arising From Parameter Changes: Population Composition

- Assume now the benchmark case but the population portion of ν of documented migrants is equal to the Malaysian nationals case i.e. 80% and the portion of ν of undocumented migrants is not the same.
- If the portion increase say from 80% to 95%, then the population of undocumented migrants decrease from the benchmark case. Lower bound here is 100%.
- If the portion decrease say from 80% to 65%, then the population of undocumented migrants increase from the benchmark case. To obtain 4 million as per UNHCR, the portion of migrants who are $\nu = v$ is 56%.

Summary of Mechanisms Driving Estimation

- The lower bound of the estimate is 1.2 million which is still a large amount of undocumented population.
- As $r_{av} \downarrow$, estimation of undocumented population decrease.
- As $pc \uparrow$, estimation of undocumented population increase.
- As population composition changes $\nu = v$ or $\nu = av$, the estimation of undocumented is ambiguous. Decrease in $\nu = v$ correlates to higher estimate of the undocumented. Increase in $\nu = v$ correlates to lower estimate of the undocumented.

Guides for Credible Estimation of Population for Biased Data

- Clarity in the data used i.e. vaccine data, internet, etc.
- Use simple models to illustrate the incentives for the subject of estimation to appear in your data.
- Produce a benchmark estimate based on the parameter values you think happens on the ground.
- Use comparative statics to show what you think happens to your estimate and then use these to identify which external estimates may be more credible than others e.g., from NGO or government bodies.

Improvements and Conclusion

- Model can be too simple but can serve as a starting point for further extensions.
- Can be used to compare different estimation values e.g., if expert estimates value under the benchmark (compositional or preference effect can explain why) or if expert estimates value above the benchmark (compositional or deterrence effect may produce the effect that outweighs preferences)
- Model is demand-driven and assumes supply of vaccines is unlimited (may be warranted given excess vaccine purchase) so it is accessible for at least one dose.