Replication Project for Development Economics: Wealth Heterogeneity and the Income Elasticity of Migration By Samuel Bazzi (2017)

Tanfei Li, Nayanika Mitash

Introduction

This replication study examines the paper 'Wealth Heterogeneity and the Income Elasticity of Migration' by Samuel Bazzi (2017), which analyzes how income shocks affect international migration from Indonesia.[1] In this paper, Bazzi addresses the opposing effects of income growth — reducing liquidity constraints while increasing opportunity costs — using a two-step estimation framework, which reveals how wealth heterogeneity shapes the migration response to income shocks. This replication seeks to unpack the methodological choices by transcribing the author's code from Stata to R and reassess the original findings in table 1, 2, 4, 5, and 6.[2]

Table 1

Table 1 summarizes key migration statistics for Indonesian villages, including: Village Population, Number of Emigrants, Emigrants/Population (migration proportion), Proportion of Villages with Emigrants Abroad, Average Number of Emigrants (in villages with migration), and Average Emigrants/Population (in villages with emigrants). These metrics provide insights into migration trends across Indonesian villages.

Stocks of International Labour Migrants

In 2005, the average Indonesian village had a population of 3,216 residents and 17 emigrants abroad. The median number of emigrants per village was 1, suggesting that most villages had few or no emigrants. The average emigration rate, calculated as the number of emigrants divided by the population, was 0.006, with a median of 0.0005. More than half of the villages (54%) had at least one emigrant. Among these villages, the average number of emigrants was 31, and the average emigration rate was 0.010.

In 2008, the average village population rose slightly to 3,377, and the average number of emigrants increased to 20. The median number of emigrants per village was 2. The average emigration rate rose to 0.007, with a median of 0.001. The proportion of villages with at least one emigrant rose to 59%. Among villages with emigrants, the average number of emigrants was 35, and the average emigration rate was 0.012.

The total number of emigrants across all villages increased from 1,113,244 in 2005 to 1,349,540 in 2008, reflecting a notable rise in international migration.

[1] Bazzi, S. (2017). Wealth heterogeneity and the income elasticity of migration. American Economic Journal: Applied Economics, 9(2), 219–255. https://doi.org/10.1257/app.20150548

[2] The replication package for all the reproduced tables below can be found in this github repository, together with the datasets and original Stata code published by the author.

Changes in International Labour Migration (2005-2008)

Between 2005 and 2008, the average increase in the number of emigrants per village was 4, though the median change was 0, indicating that many villages saw no change. The average change in the emigration rate was 0.110 percentage points, with a median change of 0. Villages with a net increase in emigrants saw an average increase of 6 individuals, and the average emigration rate change was 0.14 percentage points. Additionally, the average change in the log of the emigration rate was 0.106.

Rural vs. Urban Emigration

In both 2005 and 2008, about 59% of the Indonesian population lived in rural areas. However, a much larger proportion of emigrants originated from rural villages, with 82% of emigrants in 2005 and 83% in 2008 coming from rural areas. This indicates that international migration was more prevalent in rural parts of Indonesia.

Table 1: Summary Statistics: International Labor Migration from Indonesian Villages (Reproduced)

Stocks, 2005				
Variable	Mean	Median	SD	Max
Village population	3,216	2,095	4,123.07	78,986
Number of emigrants	17	1	60.98	1,996
Emigrants/population	0.006	0.000	0.019	0.832
1 (any emigrants abroad)	0.542	–	–	–
Number of emigrants | emigrants > 0	31	8	80.12	1,996
Emigrants/population | emigrants > 0	0.010	0.003	0.025	0.832
Stocks, 2008				
Variable	Mean	Median	SD	Max
Village population	3,377	2,187	4,330.15	82,215
Number of emigrants	20	2	64.22	998
Emigrants/population	0.007	0.001	0.020	0.759
1 (any emigrants abroad)	0.586	–	–	–
Number of emigrants | emigrants > 0	35	9	80.82	998
Emigrants/population | emigrants > 0	0.012	0.004	0.026	0.759
Changes (Δ), 2005-2008				
Variable	Mean	Median	SD	Max
Difference in number of emigrants	3.582	0.000	52.52	998
Difference in emigrants/population	0.110	0.000	1.918	59.20
Difference in emigrants (only for emigrants > 0)	6.262	1.000	73.09	995
Difference in emigrants/population (only for emigrants > 0)	0.143	0.016	2.551	59.20
Difference in ln(emigrants/population)	0.106	0.062	1.012	5.669

2005 vs 2008 National Summary		
Variable	2005	2008
Share of Indonesian population from rural areas	0.59	0.59
Share of Indonesian emigrants from rural areas	0.82	0.83
Total emigrants, all villages	1,113,244	1,349,540

Table 2

Table 2: Agricultural Income Shocks and Migration Choice in Micro Data (Reproduced)

	(1)	(2)	(3)	(4)
rainfall shock, t	0.200	0.214	1.109	1.712
	(0.104)	(0.107)	(0.586)	(0.653)
rainfall shock, t × landholdings (ha)			-0.267	-3.844
			(0.680)	(1.799)
rainfall shock, t × landholdings (ha) squared				1.583
				(0.802)
price shock, t		0.769	3.752	3.892
		(0.293)	(1.384)	(1.417)
price shock, t × landholdings (ha)			-0.581	-3.128
			(0.934)	(1.970)
price shock, t × landholdings (ha) squared				0.802
				(0.654)
Observations	1,902	1,380	1,380	1,380

Table 2 presents evidence of heterogeneous income elasticities of migration at the individual household level. The underlying model used to generate these results is as follows:

$$\Pr(migrate_{i,t+1} = 1) = \mathbf{shock}'_{vt}\boldsymbol{\alpha} + (\mathbf{shock}_{vt} \times f(land_i))'\boldsymbol{\beta} + \psi_i + \psi_t + e_{i,t+1} \tag{1}$$

The model is based on a conditional fixed effects logit framework, which estimates the probability that a household will have a member migrate abroad in the following year.

The dependent variable is a binary indicator, $migrate_{iv,t+1}$, which equals one if household i in village v experienced at least one member migrating in year $t+1$, and zero otherwise. The key independent variables include agricultural income shocks, denoted as $shock_{vt}$, and household landholding size, $land_i$.

The income shocks considered in the model consist of two components. First, the rainfall shock in year t is defined as the log deviation of the current season's rainfall from the long-run local mean. Second, the price shock in year t is measured as the log difference in the local rice price at the end of period t compared to $t-1$. To capture how the impact of income shocks on migration varies with land ownership, the model incorporates interaction terms between income shocks and a function of household landholding size, $shock_{vt} \times f(land_i)$.

To account for unobserved heterogeneity, the model includes household fixed effects, ψ_i, which control for time-invariant household characteristics that may influence migration decisions. Additionally, year fixed effects, ψ_t, are included to absorb common shocks or trends that affect all households within a given year. The error term, $e_{iv,t+1}$, captures idiosyncratic factors that may influence migration.

The coefficients estimated in this model are interpreted as average marginal effects (AMEs), which indicate the average change in the probability of migration for a unit change in the corresponding independent variable. The primary objective of the analysis is to examine how agricultural income shocks—both transitory (rainfall shocks) and more persistent (price shocks)—influence the likelihood of international migration across households with different landholdings. This provides micro-level evidence for the liquidity and opportunity cost mechanisms discussed in the paper.

Comments on key results

Income Shocks Facilitate Migration

The results in Table 2 indicate that positive agricultural income shocks increase the likelihood of international migration. The positive and precisely estimated Average Marginal Effects (AMEs) in columns 1 and 2 show that both rainfall and rice price shocks significantly raise the probability that a household sends a member abroad. This suggests that, on average, positive income shocks can help facilitate migration by alleviating financial constraints that might otherwise prevent households from covering migration costs.

Stronger Effect for Small Landholders

The inclusion of interaction terms between income shocks and household landholdings in columns 3 and 4 reveals that the income elasticity of migration is larger for smaller landholders. This finding supports the idea that liquidity constraints are more binding for poorer households. When income increases due to a positive shock, these constraints ease, making migration a viable option for households with limited land.

Non-Linear Effects and Opportunity Costs

The quadratic specification in column 4 highlights a non-linear relationship between landholdings and the effect of income shocks on migration. The AMEs for both rainfall and price shocks remain positive and significant for households with less than 0.6 hectares of land but become negative or insignificant for those with larger landholdings. This suggests the presence of both liquidity constraints (affecting smaller landholders) and opportunity cost mechanisms. Larger landowners may find migration less attractive as agricultural returns increase, leading to a weaker or even negative effect of income shocks on migration. However, the negative effect for larger landholders is not always statistically significant, likely due to data limitations such as sample size and the timing of data collection relative to price shocks.

In summary, the findings provide micro-level evidence that:

- Positive agricultural income shocks generally increase the likelihood of international migration.
- This effect is stronger for households with smaller landholdings, consistent with the alleviation of liquidity constraints.
- For households with larger landholdings, the effect of income shocks on migration is smaller and can even be negative, potentially reflecting increasing opportunity costs as agricultural income rises.

While the negative price elasticity for large landholders was not always statistically significant, the overall patterns in Table 2 suggest heterogeneous effects of income shocks on migration, depending on household wealth, proxied by land ownership.

Table 4 and 5 : Breakdown of Extensive and Intensive Responses to Agricultural Shocks

The findings from Tables 2 (and 3) provide preliminary evidence of heterogeneous migration responses to income shocks, highlighting the potential role of liquidity constraints and wealth heterogeneity. However, at the macro village level, the analysis could not disentangle whether observed changes in migration were driven by new migration flows (extensive margin) or increases in migration among existing migrant-sending villages (intensive margin).

$$E\left(\frac{M_{vs}}{N_{vs}}\right) = \underbrace{E\left(\mathbb{1}\{R_{L_s} \leq R_{iv} \leq R_{U_s}\} \mid \tilde{R}_v \geq R_{L_s}, R_v \leq R_{U_s}\right)}_{\text{intensive margin}} \times \underbrace{\Pr\left(\tilde{R}_v \geq R_{L_s}, R_v \leq R_{U_s}\right)}_{\text{extensive margin}} \quad (2)$$

The extensive margin of migration refers to whether a village has any international migrants at all, capturing the decision to transition from zero to positive emigration. This depends on two conditions: at least one individual must be able to afford migration—determined by wealth exceeding a threshold in the presence of cash-in-advance (CIA) constraints—and at least one individual must find migration profitable, with expected net returns exceeding home income. Traditional random utility models struggle to explain the prevalence of zero migration, as they assume continuous migration flows. In contrast, the intensive margin describes the scale of migration within villages already participating in migration, influenced by expected income gains relative to staying home. Income shocks, such as rainfall deviations and rice price changes, interact with individual wealth heterogeneity to determine the rate of migration, with liquidity constraints shaping the responsiveness of migration flows.

In the following section, the author employs a two-step Heckman-style selection framework to disentangle these two margins to provide a more accurate estimation of the income elasticity of migration. Notably, this approach manages to: (1) explicitly model the *extensive margin* by addressing the prevalence of zero migration flows, an issue ignored in previous reduced-form estimates which can lead to underestimation of income elasticity of migration; (2) correct for *selection bias* by accounting for the fact that migration flows are only observed in villages with existing migration, leading to more accurate estimates of income elasticity; and (3) provide a clearer understanding of how *wealth heterogeneity interacts with income shocks*, influencing both the decision to initiate migration and the scale of ongoing migration.

Table 4

Table 4: Extensive Margin First-Stage Estimates (Reproduced)

	SU-LPM				Bivariate probit			
	(1)		(2)		(3)		(4)	
Variable	2008	2005	2008	2005	2008	2005	2008	2005
log maximum landholdings	0.032 (0.002)	0.035 (0.002)			0.123 (0.008)	0.126 (0.008)		
log minimum landholdings	-0.045 (0.005)	-0.035 (0.005)			-0.180 (0.019)	-0.140 (0.019)		
Pareto exponent $\widehat{\lambda}$			-0.006 (0.002)	-0.010 (0.003)			-0.037 (0.012)	-0.054 (0.012)
log village population			0.076 (0.003)	0.071 (0.003)			0.279 (0.013)	0.253 (0.012)
log district population less v	0.154 (0.005)	0.141 (0.005)	0.108 (0.005)	0.101 (0.005)	0.541 (0.022)	0.496 (0.021)	0.373 (0.024)	0.344 (0.023)
log district area less v	-0.059 (0.003)	-0.062 (0.003)	-0.058 (0.004)	-0.060 (0.003)	-0.182 (0.013)	-0.193 (0.012)	-0.185 (0.013)	-0.195 (0.012)
log distance to subdistrict capital	-0.021 (0.003)	-0.023 (0.003)	-0.019 (0.003)	-0.021 (0.003)	-0.083 (0.010)	-0.084 (0.010)	-0.076 (0.010)	-0.077 (0.010)
log distance to nearest emigration center	-0.019 (0.005)	-0.012 (0.005)	-0.019 (0.004)	-0.013 (0.005)	0.003 (0.019)	0.009 (0.018)	0.003 (0.019)	0.011 (0.018)
rice price shock	0.091 (0.082)	0.237 (0.102)	0.082 (0.076)	0.209 (0.089)	0.145 (0.391)	0.954 (0.373)	0.151 (0.393)	0.875 (0.374)
rainfall shock	0.034 (0.004)	0.034 (0.004)	0.034 (0.004)	0.034 (0.004)	0.094 (0.019)	0.093 (0.017)	0.101 (0.019)	0.100 (0.017)
Number of Villages	44665		44665		44665		44665	

Notes: All replicated coefficients are close to the author's results. The standard errors in our replication procedure are **not** clustered at the district level because we could not find a clustring procedure that is compatible with the the system function for SU-REG and the biprobit function in R, whereas clustering SE is built-in in Stata.

Table 4 of the paper focuses on the **extensive margin of migration**. The author estimates the probability of observing any migrants from a village in 2005 and 2008 using two main estimation techniques:

Column 1 & 2: Seemingly Unrelated Linear Probability Model (SU-LPM)

This method jointly estimates linear probability models for the presence of migrants in 2005 and 2008, accounting for potential correlations in the error terms across the two periods. The coefficients in a linear probability model are interpreted as the change in the probability of the outcome (having any migrants) for a one-unit change in the independent variable. The standard errors for the SU-LPM estimates are obtained using a block bootstrap procedure clustered at the district level.

Column 3 & 4: Bivariate Probit Estimator

This method models the binary outcome of having any migrants in 2005 and 2008 using a bivariate probit model, assuming that the unobserved factors influencing migration in the two periods follow a bivariate normal distribution. The table reports the coefficients from these probit regressions. In probit models, the coefficients represent the change in the z-score of the probability of the outcome, while the actual change in probability (marginal effects) is typically derived from these coefficients.

Estimation Equation

The general form of the first-stage equation estimated is:

$$\Pr(M_{vs} > 0) = \mathbf{shock}'_{vs}\eta_{\mathbf{s}} + \mathbf{f}(\tilde{\mathbf{R}}_{\mathbf{v}}, \mathbf{R}_{\mathbf{v}})'\rho_{\mathbf{s}} + \mathcal{Z}'_{\mathbf{vs}}\gamma_{\mathbf{s}} + \mathbf{u}_{\mathbf{vs}}, \tag{3}$$

where $\Pr(M_{vs} > 0)$ is the probability that village v has international migrants at time s (2005 or 2008). $\mathbf{shock}_{vs}$ represents rainfall and price shocks for year s. $f(\tilde{R}_v, \bar{R}_v)$ is a function of the maximum ($\tilde{R}_v$) and minimum ($\bar{R}_v$) landholdings; in column 2, it is replaced by the Pareto exponent ($\hat{\lambda}_v$) and log village population. $\mathcal{Z}_{vs}$ is a vector of control variables and excluded instruments, including log distance to the subdistrict capital and nearest emigration center, share of households with landholdings above 0.1 hectares, share of wetland in total agricultural land, urban classification indicator, Muslim, ethnic Chinese, and ethnic Arab population shares, and accessibility by motorized land transport. $\eta_{\mathbf{s}}, \rho_{\mathbf{s}}, \gamma_{\mathbf{s}}$ are coefficient vectors to be estimated, and u_{vs} is the error term.

Comments on key results

Column 1 & 2

landholdings

The positive coefficient on log maximum landholdings suggests that villages with wealthier households are more likely to have migrants. This supports the idea that some level of wealth is necessary to overcome migration costs. The negative coefficient on log minimum landholdings suggests that villages where even the poorest have relatively large landholdings are less likely to have migrants, likely due to lower migration incentives.

Pareto exponent $\hat{\lambda}_v$

The negative coefficient on the Pareto exponent ($\hat{\lambda}_v$) suggests that villages with a higher upper tail of the landholding distribution are more likely to have migrants, since the exponent is inversely proportional to the share of land-rich residents in a village. The positive coefficient on log of village population indicates that larger villages are more likely to have international migrants due to a higher likelihood of having individuals with financial means to migrate, as implied by the Pareto distribution of wealth.

Columns 2 & 4

District-Level Variables

The positive coefficient on log district population and the negative coefficient on log district area suggest that migration recruiters are more likely to be active in densely populated districts, lowering migration costs. The negative coefficients on distance to subdistrict capital and emigration center suggest that proximity to administrative and migration hubs increases the likelihood of migration.

Rainfall and Rice Price Shocks

The coefficients on rainfall and rice price shocks are generally positive but statistically insignificant, indicating that these shocks impact income but may not be sufficient to push villages over the migration threshold. Notably, both rainfall and rice price shocks are included as they represent distinct types of income fluctuations – the former temporary and the later more persistent – allowing the author to rigorously test the theoretical predictions of the model regarding the roles of liquidity constraints and opportunity costs in shaping the income elasticity of migration

Comments on underlying assumptions

Critically, the assumption of exogeneity for variables like landholdings and population might be challenged as migration itself could influence local economies and potentially long-term land distribution or population dynamics. The bivariate normality assumption in the probit model regarding unobserved factors is a strong assumption that may not accurately reflect the underlying complexities. Furthermore, relying solely on a log-linear functional form for the relationship between landholdings and migration probability might oversimplify potentially non-linear effects. Finally, while landholdings are likely a key wealth indicator in rural Indonesia, they may not fully capture the nuances of household wealth, potentially overlooking other relevant assets or access to credit that could also influence the affordability of migration.

Table 5

Two-Step Estimation Framework

The methodology for Table 5 employs a two-step estimation framework to account for villages with zero international migration in the observed periods. This approach separates the decision of whether a village has any migrants (the extensive margin) from the effect of income shocks on the scale of migration in villages already sending migrants (the intensive margin).

Step 1: Estimating the Extensive Margin (First Stage)

The first step, as presented in Table 4, estimates the probability that a village has any international migrants in 2005 and 2008 using the following specification:

This first-stage estimation is conducted using a Seemingly Unrelated Linear Probability Model (SU-LPM) and a bivariate probit estimator to account for the binary nature of the dependent variable and panel structure of the data.

Step 2: Estimating the Intensive Margin (Second Stage)

The second step, presented in Table 5, examines the change in the log migration rate between 2005 (t) and 2008 ($t+1$) for villages with migrants in both periods. The estimating equation is:

$$\begin{aligned}\Delta \ln\left(\frac{M_{v,t+1}}{N_{v,t+1}}\right) &= \theta_a \,\Delta\text{rainfall shock}_{vt} + \theta_{a\lambda}\left(\Delta\text{rainfall shock}_{vt} \times \hat{\lambda}_v\right) \\ &\quad + \theta_p \,\Delta\text{price shock}_{vt} + \theta_{p\lambda}\left(\Delta\text{price shock}_{vt} \times \hat{\lambda}_v\right) \\ &\quad + \mathbf{X}_v'\boldsymbol{\theta} + \mathcal{D}_{j(v)} + f\left(\hat{P}_{v,t+1}, \hat{P}_{vt}\right) + \Delta\varepsilon_{v,t+1} \qquad (4)\end{aligned}$$

where $\Delta \ln \left(\frac{M_{v,t+1}}{N_{v,t+1}} \right)$ represents the change in the log migration rate (number of emigrants divided by village population) from 2005 to 2008. Δ*rainfall shock*$_{vt}$ and Δ*price shock*$_{vt}$ measure the changes in cumulative rainfall and annualized rice price shocks between the two periods. $\hat{\lambda}_v$ is the estimated Pareto dispersion parameter for paddy landholdings, capturing wealth heterogeneity. The interaction terms allow income shock effects to vary with village wealth distribution. Higher $\hat{\lambda}_v$ signifies less dispersion and a higher concentration of small landholders. $\mathbf{X}_v$ includes time-invariant controls such as $\hat{\lambda}_v$ and migration cost proxies (e.g., distance to emigration points, ethnic group composition, plurality destination in 2005). $\mathcal{D}_{j(v)}$ represents fixed effects for the plurality migration destination of village v in 2005, controlling for destination-specific factors. $f(\hat{P}_{v,t+1}, \hat{P}_{vt})$ are selection correction terms from the first-stage estimation, which address potential bias from observing migration changes only in villages with migrants in both periods. These can take the form of bivariate Mills ratios (in the parametric Poirier method) or polynomials in propensity scores (in the semi-parametric Das, Newey, and Vella (DNV) method). The joint significance of these terms, reported in Table 5, indicates the importance of selection correction. $\Delta\varepsilon_{v,t+1}$ is the error term.

The second-stage estimation is performed using OLS (without selection correction in column 1), and with both semi-parametric (DNV) and parametric (Poirier) methods to correct for selection bias. Standard errors in Table 5 are clustered at the district level, with significance levels for the selection-corrected estimates obtained via block bootstrap-t procedures to account for potential heteroskedasticity and intra-district correlation.

This methodology applies a two-step approach: the first stage determines whether a village participates in international migration, while the second examines how income shocks influence migration scale in villages already engaged in migration. Selection correction terms from the first-stage estimation address potential biases. The inclusion of interaction terms with the Pareto dispersion parameter allows for estimating heterogeneous income elasticities of migration based on wealth distribution. A key assumption underlying this methodology is the validity of exclusion restrictions in the first stage, ensuring that selection correction terms properly account for bias in the second-stage estimates.

Table 5: Two-Step Estimates of the Income Elasticity of Migration (Reproduced)

	OLS				
Panel A. Semiparametric correction procedure	(1)	(2)	(3)	(4)	(5)
Δ rainfall shock	0.077	0.293	0.143	0.118	-0.029
	(0.133)	(0.177)	(0.076)	(0.081)	(0.187)
Δ price shock	-0.078	0.337	0.341	0.886	-2.225
	(0.443)	(0.584)	(0.635)	(0.492)	(0.836)
Δ rainfall shock $\times \widehat{\lambda}$			0.053	0.097	0.085
			(0.227)	(0.231)	(0.058)
Δ price shock $\times \widehat{\lambda}$				-1.214	0.823
				(1.007)	(0.390)
Δ rainfall shock $\times$ share households $>$ 0.1 Ha					0.617
					(0.148)
Δ price shock $\times$ share households $>$ 0.1 Ha					4.069
					(0.907)
Joint significance of selection correction terms		***	***	***	***

Panel B. Parametric correction procedure	-	(6)	(7)	(8)	(9)
Δ rainfall shock		0.244	0.034	0.099	0.028
		(0.137)	(0.174)	(0.178)	(0.190)
Δ price shock		0.518	0.525	-1.610	-1.907
		(0.457)	(0.457)	(0.783)	(0.839)
Δ rainfall shock × $\widehat{\lambda}$			0.128	0.088	0.037
			(0.057)	(0.061)	(0.062)
Δ price shock × $\widehat{\lambda}$				1.265	0.917
				(0.404)	(0.378)
Δ rainfall shock × share households > 0.1 Ha					0.399
					(0.142)
Δ price shock × share households > 0.1 Ha					2.091
					(0.793)
Joint significance of selection correction terms		***	***	***	***
Number of villages	24855	24855	24855	24855	24855

Note: Column 1 shows estimates that deviate slightly from the author's results, while all other columns present close replicates for both the estimated coefficients and the standard error clustered at the district level (compatible solution found for the felm function in R, which is used for the second-stage regression). The second-stage linear programme is run on villages with positive (ie. non-zero) migration, thus dropping the number of valid observations to 24855. The bootstrap was run at district level with 1000 replications for Panel A but 500 for Panel B as Panel B is computationally heavy. The selection correction terms are calculated from the fitted values in the first-stage regression in each bootstrap iteration: in Panel A, they are calculated as fitted values and their interactions up to the 3rd polynomial term in each iteration, while Panel B constructs them with formulae of Inverse Mill Ratio proposed by Poirier(1980). The joint significance test *** denotes P-values < 0.001, which again matches the author's results.

Comments on key results

The results from Table 5 indicate that, after accounting for selection into migration, agricultural income shocks—both rainfall and rice price shocks—positively influence the change in village-level migration rates. This suggests that for villages already engaged in migration, liquidity constraints play a crucial role in shaping responses to income fluctuations. Additionally, the effect of income shocks on migration is not uniform across villages; those with a higher concentration of small landholders (higher $\hat{\lambda}_v$, indicating less dispersion) experience larger increases in migration rates following both types of income shocks. This supports the hypothesis that liquidity constraints are more binding for poorer households, making their migration decisions more sensitive to economic shocks.

Moreover, failing to account for the extensive margin of migration, as seen in the OLS estimates in column 1, leads to an underestimation of financial constraints' significance in migration decisions. The selection correction terms are jointly significant, demonstrating that the same factors determining whether a village participates in migration also influence the scale of migration. These findings underscore the importance of using selection-corrected models to accurately capture the relationship between income fluctuations and migration responses.

Although I managed to closely reproduce the author's results, I also noticed that it was difficult for the model to converge when running on the parametric model. The statistic exhibits sensitivity to variations in the sample, leading to inconsistencies in the bootstrap estimates. This may indicate the high precision and sensitivity of the parametric model as a good check for the semi-parametric results. However, it might also be evidence that there might be clustering effects or district-level heterogeneity that violates the parametric assumptions.

Table 6

Table 6 extends the two-step estimation framework used in Table 5 to further explore the mechanisms underlying the income elasticity of migration, particularly the role of opportunity costs. The dependent variable remains the change in the log migration rate between 2005 and 2008, consistent with the second stage of the previous model. The semi-parametric (DNV) selection correction procedures employed in Table 5 is applied here to account for the extensive margin of migration.

Table 6 introduces interaction terms to examine heterogeneous wealth effects. Column 1 introduces interaction terms between both rainfall and price shocks and an indicator for the presence of recruitment agencie. This accounts for the role of migration intermediaries in shaping the effects of income shocks. Column 2 introduces interactions between income shocks and the quartile of the district's agricultural GDP in 2002, allowing for variation in income shock effects based on regional agricultural development. The regressions also include the same vector of time-invariant control variables as in Table 5, such as the Pareto exponent $\hat{\lambda}_v$ and proxies for migration costs, along with fixed effects for the plurality migration destination in 2005. Standard errors are clustered at the district level, and significance levels are determined using a block bootstrap-t procedure.

Comments on key results

Recruitment Agencies and Opportunity Costs: Price shocks have a strong positive effect on migration in villages without recruitment agencies, indicating that liquidity constraints limit migration in these areas, and increased income from price shocks facilitates migration. In contrast, price shocks have no significant effect on migration in villages with recruitment agencies, suggesting that migration is more accessible in these areas due to interlinked contracts or lower upfront costs. If liquidity constraints were the only factor, price shocks should have a similar effect to rainfall shocks. Rainfall shocks positively affect migration in both types of villages, but the effect is significantly larger in villages without recruitment agencies, reinforcing the argument that cash-in-advance constraints are more binding where recruiters are absent. The asymmetric response to transitory (rainfall) and persistent (price) shocks in villages with recruiters highlights the role of opportunity costs.

Agricultural Development and Opportunity Costs: In the lowest quartile of agricultural GDP, both rainfall and price shocks significantly increase migration, demonstrating that liquidity constraints are the dominant factor in the poorest agricultural areas. However, in the highest quartile of agricultural GDP, rainfall shocks have no effect, while price shocks negatively affect migration at the 5 percent significance level. This suggests that in more developed agricultural regions, the opportunity costs of migration outweigh liquidity constraints, as rising rice prices incentivize staying and expanding farm production. The second and third quartiles display a transition, where the positive effect of income shocks on migration diminishes as agricultural GDP increases, reflecting a shift from liquidity-constrained to opportunity-cost-driven migration decisions.

Table 6 provides compelling evidence that the relationship between income shocks and migration is not uniform. Liquidity constraints are a dominant factor in poorer areas and those without established migration networks, where increased income facilitates migration. However, in more developed agricultural regions with potentially lower migration barriers, the opportunity cost of leaving becomes increasingly important, leading to a reduction in migration in response to persistent positive income shocks like sustained rice price increases.

Table 6: Evidence on the Opportunity Cost Mechanism (Reproduced)

	(1)	(2)
Δ rainfall shock × recruiter presence	0.245	
	(0.156)	
Δ rainfall shock × no recruiter presence	0.442	
	(0.135)	
Δ price shock × recruiter presence	0.138	
	(0.628)	
Δ price shock × no recruiter presence	1.780	
	(0.473)	
Δ rainfall shock × agricultural GDP, quartile = 1		0.570
		(0.304)
Δ rainfall shock × agricultural GDP, quartile = 2		0.183
		(0.202)
Δ rainfall shock × agricultural GDP, quartile = 3		0.217
		(0.188)
Δ rainfall shock × agricultural GDP, quartile = 4		-0.055
		(0.164)
Δ price shock × agricultural GDP, quartile = 1		3.352
		(1.075)
Δ price shock × agricultural GDP, quartile = 2		0.680
		(0.552)
Δ price shock × agricultural GDP, quartile = 3		-0.995
		(0.764)
Δ price shock × agricultural GDP, quartile = 4		-1.970
		(0.691)
Number of villages	24855	24493

Notes: All replicated coefficients and standard errors are close to the author's results. The standard errors are properly clustered at the district level. The observation counts also follow the original results.

Conclusion

This replication study reinforces the key insights of the original paper, confirming the importance of both liquidity constraints and opportunity costs in shaping international migration flows. By replicating the two-step econometric framework in R, this analysis validates the findings that positive income shocks increase migration in liquidity-constrained villages, particularly those with a greater proportion of small landholders. Additionally, the results support the notion that in more developed agricultural regions and areas with established migration networks, opportunity costs become more significant, potentially dampening migration responses to persistent income growth. This replication further highlights the structural role of migration costs and the prevalence of liquidity constraints, emphasizing the potential for policy interventions to facilitate mobility. It was a challenging yet rewarding exercise for our course on development economics.