1 Estimate remittances sent (in % of income)

$$\rho_{od} = \delta_1 ln(ypc_d) + \delta_2 ln(\frac{ypc_d}{ypc_o}) + \delta_3 \phi_{od} + \epsilon_{od}$$
$$ln(\rho_{od}) = \delta_1 ln(ypc_d) + \delta_2 ln(\frac{ypc_d}{ypc_o}) + \delta_3 \phi_{od} + \epsilon_{od} \quad (1)$$

Estimation at country*country level. Using data on bilateral remittance flows and costs from the World Bank for 2017 (only data available for remittances):

	remshare		ln(remshare)	
	(1)	(2)	(3)	(4)
ypc_o	-0.020**		-0.241**	
	(0.006)		(0.081)	
ypc_d	0.001	-0.019*	-0.362***	-0.603***
	(0.004)	(0.009)	(0.031)	(0.090)
ypc ratio		0.020**		0.241**
		(0.006)		(0.081)
remcost (ϕ)	1.276*	1.276*	-5.953	-5.953
	(0.616)	(0.616)	(3.642)	(3.642)
(Intercept)	0.154*	0.154*	3.418***	3.418***
	(0.069)	(0.069)	(0.831)	(0.831)
Estimator	OLS	OLS	OLS	OLS
N	34,225	34,225	10,442	10,442
R^2	0.005	0.005	0.133	0.133

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 1: Estimation of remittances sent as share of migrant's income. Results from OLS regression of remittance shares (columns 1–2) and log remittance shares (columns 3–4). Dependent variables are regressed on income per capita at destination and cost of sending remittances (all columns), as well as income per capita at origin (columns 1 and 3) and ratios of destination to origin incomes (columns 2 and 4). Standard errors are clustered at the origin and destination levels.

Estimation at region*region level. Using the following aggregation strategies:

- For remittance cost ϕ: weighting corridors by remittance flows
- For remittance share ρ: weighting corridors with weight w_{ij}:

$$\forall (i,j) \in (o,d) \quad w_{ij} = \frac{stock_{ij}}{\sum_{i,j} stock_{ij}} * \frac{ypc_{mig}}{\frac{\sum_i ypc_{mig} * pop_i}{\sum_i pop_i}} \quad (2)$$

	remshare reg		ln(remshare reg)	
	(1)	(2)	(3)	(4)
$ypc_{o,reg}$	-0.965**		-4.957***	
	(0.306)		(0.878)	
$ypc_{d,reg}$	-0.573***	-1.538***	-1.579*	-6.536***
	(0.102)	(0.423)	(0.650)	(0.739)
ypc ratio		0.965***		4.957***
		(0.100)		(0.878)
remcost (ϕ_{reg})	2.035*	2.035*	9.399	9.399
	(0.790)	(0.790)	(6.254)	(6.254)
(Intercept)	3.490***	3.490***	11.799***	11.799***
	(0.295)	(0.104)	(1.609)	(1.609)
Estimator	OLS	OLS	OLS	OLS
N	256	256	251	251
R^2	0.316	0.316	0.411	0.411

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 2: See legend in Table 1.

Comparison of residuals from the two methods. We compare residuals from equation 1, more precisely $exp(\epsilon_{od})$, using:

- Direct residuals of the region*region level estimation
- Aggregated residuals of the country*country level estimation by weighting $exp(\epsilon_{ij})$ with weigths w_{ij} defined in equation 2

A comparison of residuals from both methods of estimation is displayed in Figure 1.

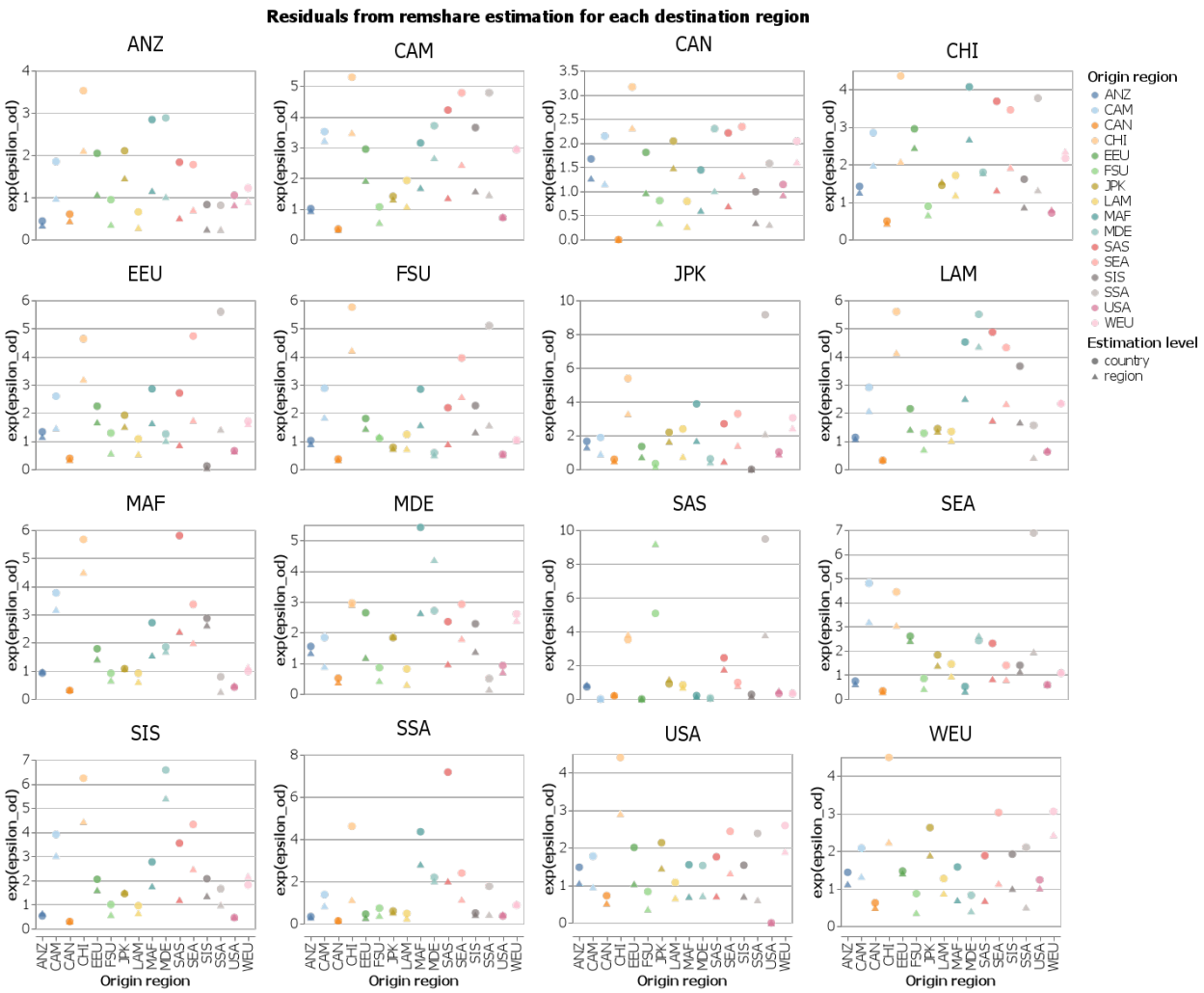

Figure 1: Residuals of equation1 estimating the share of income that a migrant sends back home as remittance. Residuals estimated at the country*country (circles) and region*region (triangles) level. Each of the 16 panels illustrates a destination region. The color code illustrates origin regions.

2 Estimate gravity model considering remittance shares as endogenous

$$ln(move_{od}) = ln(\beta_0) + \beta_1 ln(pop_o(t)) + \beta_2 ln(pop_d(t)) + \beta_4 ln(ypc_o(t)) + \beta_5 ln(ypc_d(t)) + \beta_7 ln(dist_{od}) + \beta_8 exp(\epsilon_{od}) + \beta_9 \phi_{od} + \beta_{10} \psi_{od} + (\alpha_o + \gamma_d) + \delta_t + \varepsilon_{odt} \quad (3)$$

Estimation at country*country level. Using data on bilateral migration flows from Azose and Raftery (2018) as computed in Abel and Cohen (2019) for 1990–2015, in 5-year periods; as well as residuals ϵ_{od} from equation 1 using specification in column (4) of Table 1:

	migration flow	
	(1)	(2)
pop_o	0.689***	0.709**
	(0.040)	(0.240)
pop_d	0.686***	-0.722**
	(0.042)	(0.238)
ypc_o	0.417***	0.169
	(0.060)	(0.136)
ypc_d	0.830***	0.010
	(0.070)	(0.096)
distance	-1.297***	-1.487***
	(0.063)	(0.064)
$exp(\epsilon_{od})$	0.011*	0.010*
	(0.005)	(0.005)
remcost (ϕ)	-9.670	-12.727
	(15.655)	(15.133)
comofflang (ψ)	1.743***	1.591***
	(0.133)	(0.130)
Year FE	Yes	Yes
Origin FE		Yes
Destination FE		Yes
Estimator	OLS	OLS
N	73,397	73,397
R^2	0.479	0.611
Within-R^2	0.478	0.366

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 3: Estimation of gravity model. Results from OLS regression of migration flows. Specifications with year fixed effects (all columns), and origin and destination fixed effects (column 2). Estimates are obtained based on data from Azose and Raftery (2018) as computed in Abel and Cohen (2019). Intercept computed as constant plus average values of year fixed effects. Standard errors are clustered at the origin and destination levels.

Estimation at region*region level. Using the same data, and the following aggregation strategies:

- For distances *dist*: distances between centers of population of regions, computed as aryhtmetic means of coordinates of a region's countries' capitals weighted by population
- Direct residuals of the region*region level estimation from equation 1 using specification in column (4) of Table 2
- For remittance cost ϕ: weighting corridors by remittance flows
- For common offficial language ψ: weighting corridors by migrant flows

	migration flow	
	(1)	(2)
pop_o	0.676	1.226
	(0.879)	(0.935)
pop_d	0.524	1.581***
	(2.767)	(0.449)
ypc_o	0.895	0.429
	(0.512)	(0.371)
ypc_d	1.427	0.473
	(1.119)	(1.279)
distance	-0.711	-0.680
	(4.348)	(0.562)
$exp(\epsilon_{od})$	-0.293	-0.190
	(1.364)	(0.278)
remcost (ϕ)	14.647	-1.379
	(8.950)	(3.933)
comofflang (ψ)	2.183***	0.504
	(0.322)	(0.461)
(Intercept)	-32.122	-0.008
Year FE	Yes	Yes
Origin FE		Yes
Destination FE		Yes
Estimator	OLS	OLS
N	1,170	1,170
R^2	0.559	0.797
Within-R^2	0.556	0.471

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 4: See legend in Table 3.

Comparison of residuals from the two methods. We compare residuals from equation 3, more precisely $move_{od} - \hat{f}_{grav}(...)$. A comparison of residuals from both methods of estimation is displayed in Figure 2.

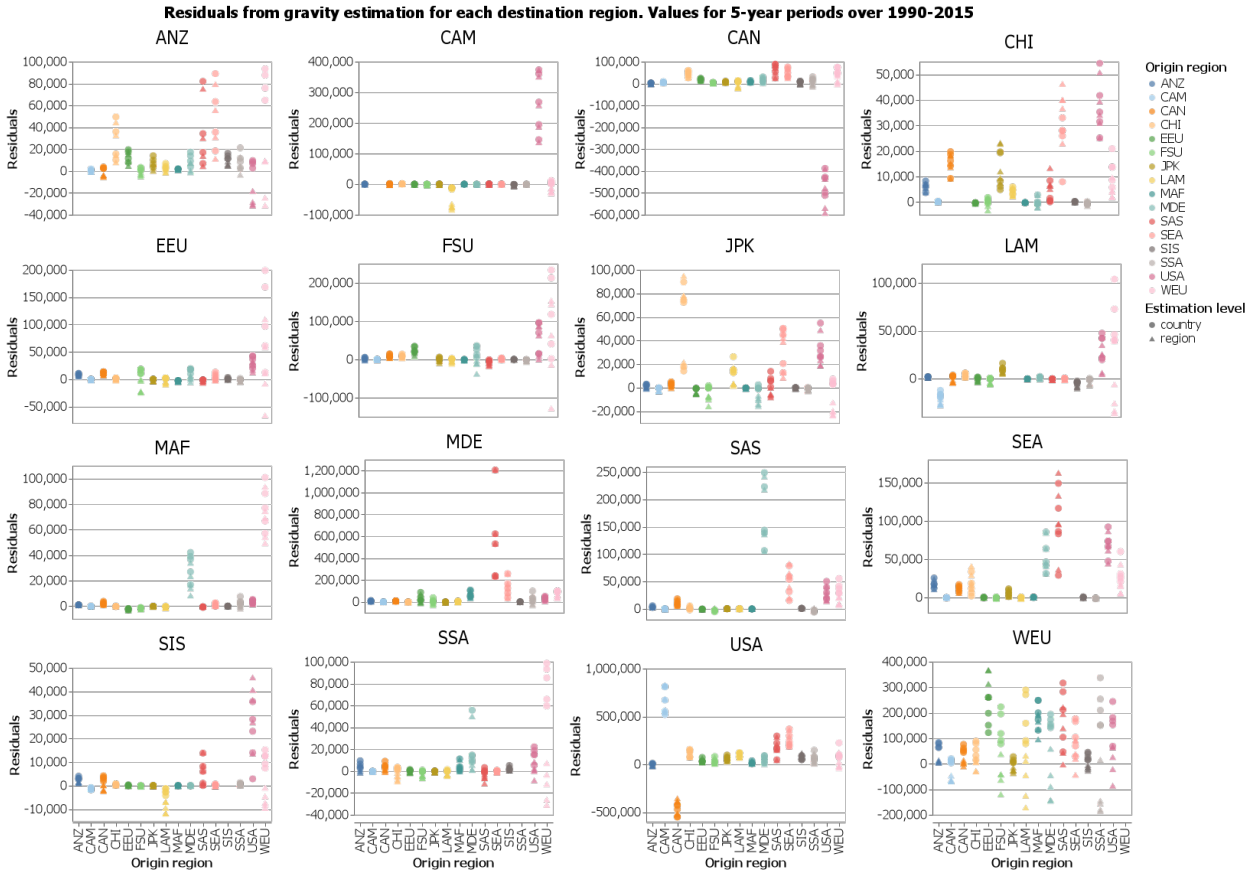

Figure 2: Residuals of equation 3 estimating the number of people moving between two regions following our gravity model. Residuals estimated at the country*country (circles) and region*region (triangles) level. Each of the 16 panels illustrates a destination region. The color code illustrates origin regions. Values from five 5-year periods, spanning 1990-2015, are presented.