The Effect of Internal Migration on Crime and Violence: Evidence from Indonesia

Brian Feld[§] and Marieke Kleemans[¥]

[§]Universidad EAFIT
[¥]University of Illinois, Urbana-Champaign

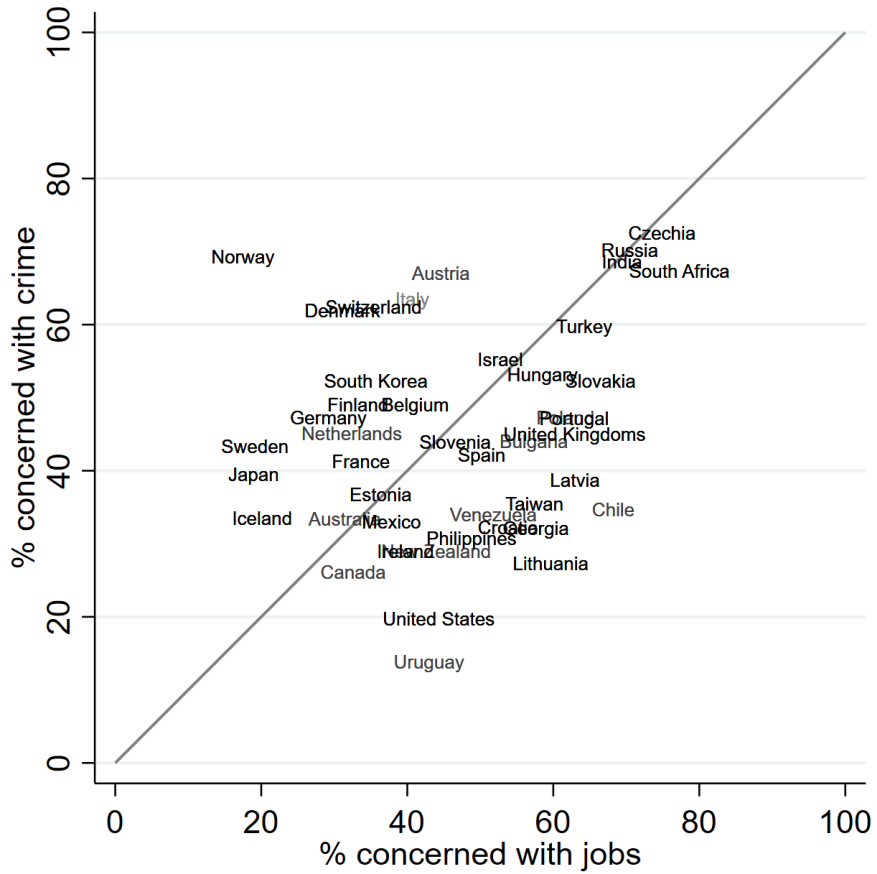
October 2022

Motivation

- Public debate on migration with concerns that immigrants
  - Worsen labor market outcomes at destinations
  - Increase crime at the destination
- Labor market impact of immigration well-studied
- Smaller literature on the impact of immigration mostly focussed on international migration into high-income countries (Bianchi et al, 2012; Bell et al, 2013; Jaitman and Machin, 2013; Chalfin, 2014, and Spenkuch, 2014; Ozden et al., 2018; Egger, 2021)
- Small increase only in economic crimes, some find no effect

Public opinion vs empirical evidence

- Discrepancy between public opinion and empirical evidence
- Agreement with statements "Immigrants increase crime rates" and "Immigrants take jobs away from natives" (ISSP, 2013)



This paper

- Research question: What is the causal effect of internal migration on crime and violence in Indonesia?
  - Weaker institutions and higher crime rates
  - Large informal sector could absorb excess labor supply
  - Lower cost of internal migration implies larger flows and different selection
- We combine a 10 year panel of 32,000 individuals migrating within Indonesia with crime reports from 2 million newspapers articles and survey data on crime victims
- To address endogeneity of migration decision, we instrument migration at destination with rainfall shocks at origin interaction with historical migration patterns, following Kleemans and Magruder (2018)
  - 1 pp increase in the share of migrants ($\approx$ 6% increase) $\rightarrow$ 0.97% decrease in income and 0.24 pp decrease in employment

Empirical Strategy

- Our interest lies in studying the relationship between crime rates and migration:

$$crime_{dt} = \beta_0 + \beta_1 migrant_{dt} + \gamma X_{dt} + \delta_t + \alpha_d + \varepsilon_{dt}$$

$(d = destination, \quad t = time)$

- $crime_{dt}$ is the crime rate per 100,000 people
- $migrant_{dt}$ is the percentage of migrants
- Identification of β_1 relies on the assumption that the choice whether and where to migrate is not correlated with the error term
  - OLS estimates will be biased if omitted variables affect both crime and migration
  - Direction of bias ex-ante unknown

Predictions

Empirical Strategy

- First stage: Migrants' responsiveness to weather shocks (Munshi, 2003) following Kleemans and Magruder (2018)

$$migrant_{dt} = b \sum_{o \epsilon C(d)} (w_o rainfall_{o,t-1}) + cX_{dt} + d_t + a_d + e_{dt}$$

- $\sum_{o \in C(d)} (w_o rainfall_{o,t-1})$ = rainfall in origin area o at time $t-1$, summed over all origin areas of catchment area $C(d)$ with weights w_o proportional to the number of migrants that historically come from that origin area
- Second stage: Effect of rainfall-induced migration on crime at destination

$$crime_{dt} = \beta \widehat{migrants}_{dt} + \gamma X_{dt} + \delta_t + \alpha_d + \varepsilon_{dt}$$

- Cluster at destination level

Identification

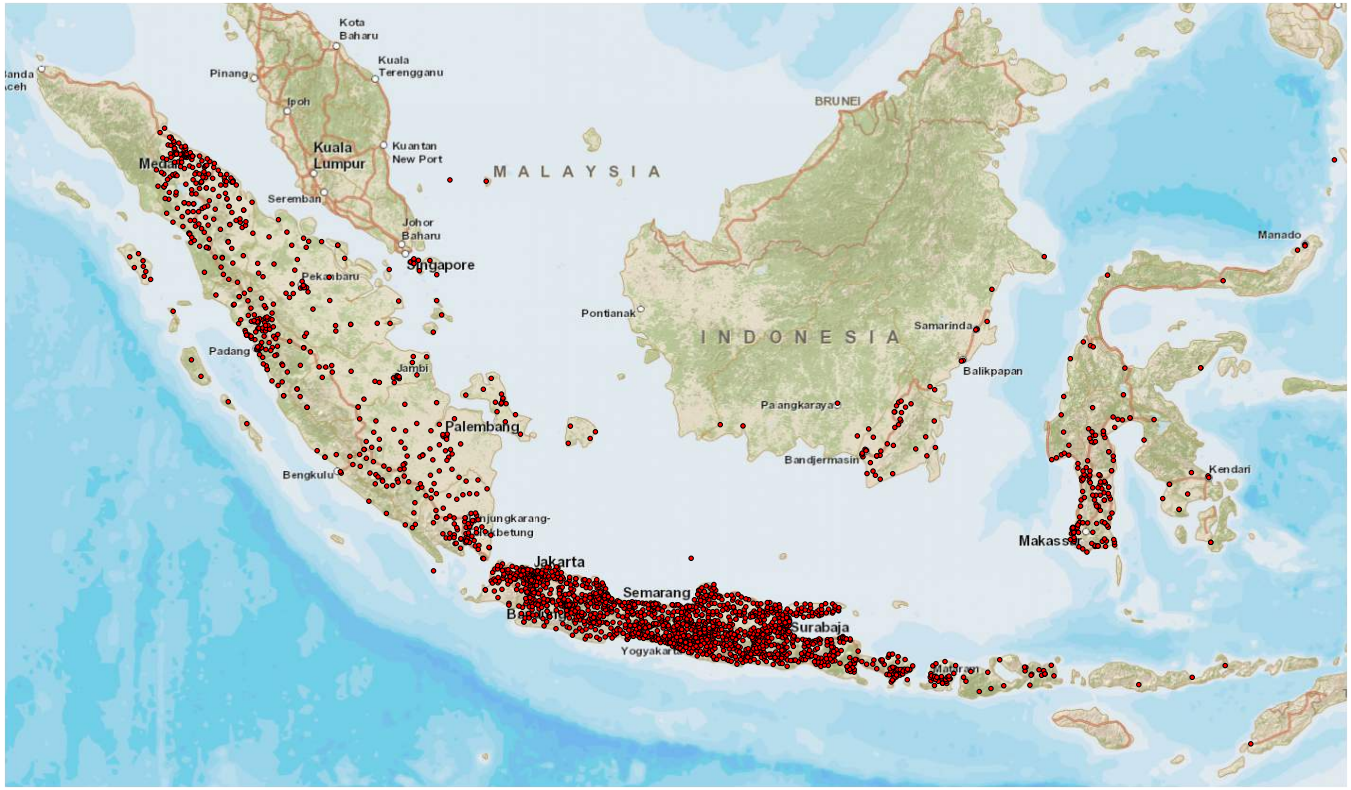
- Serial correlation in weather variable
  - Control for weather at the destination
- Identifying assumption that precipitation in migrant origin areas only affects crime and violence at migrant destinations through migration from origin to destination, once precipitation in destination areas is controlled for
- Local average treatment effect of rainfall-induced migration
- Short term impacts

Testing the Exclusion Restriction

Migration data

- Indonesia Family Life Survey (IFLS) 5 waves
  - Representative of 83% of Indonesian population (total pop. 270M)
  - Less than 5% attrition across waves
- Longitudinal data for 32,701 individuals, period 2005 - 2014
  - Current and retrospective annual data on individuals' locations
  - Migrant defined as person who does not live in their place of birth
  - All moves at least 6 months are included, shorter/seasonal moves excluded
  - Locations are defined at the sub-district level (kecamatan)
  - Mostly internal migration: $> 99\%$
- Migrant share: Share of migrants in each destination-year

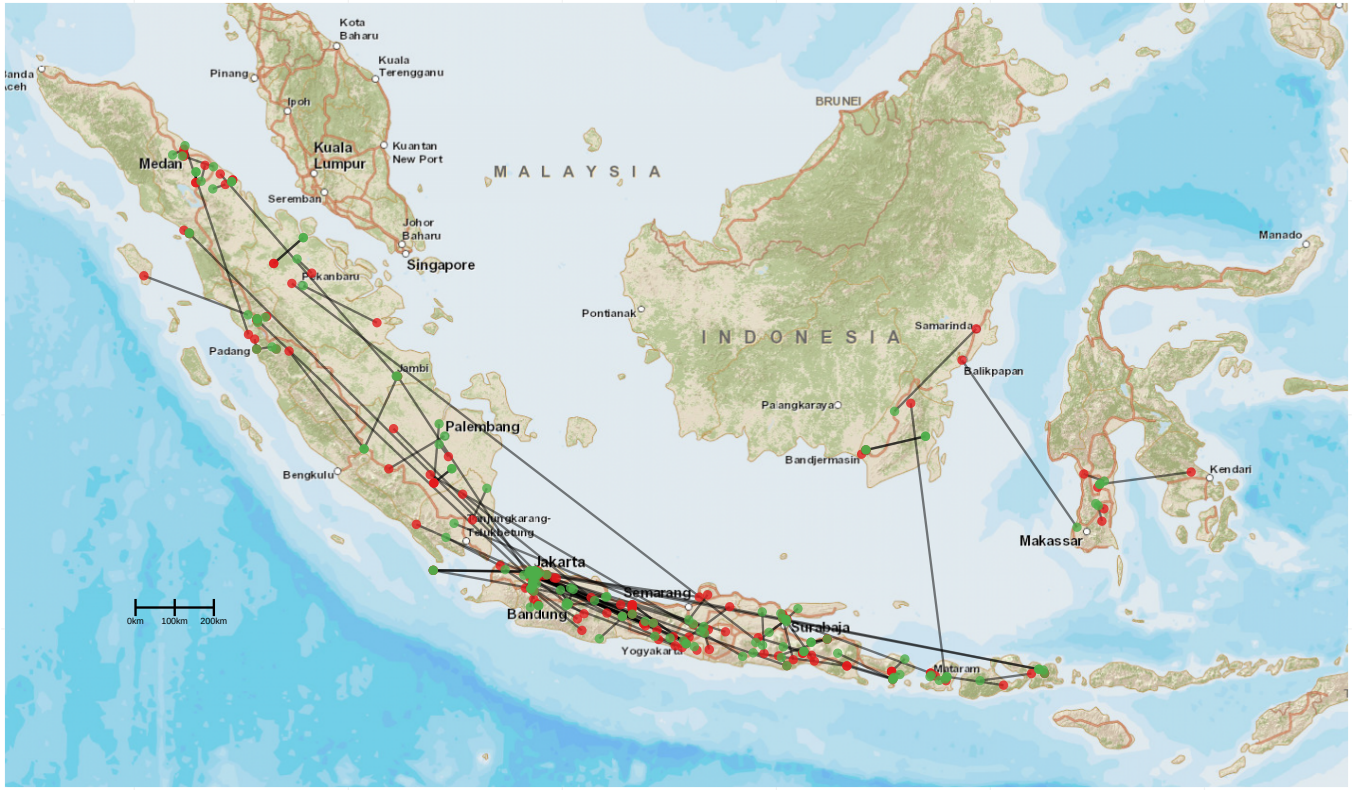
Migration data



Panel

Migration data

Year: 1995 Month: 08



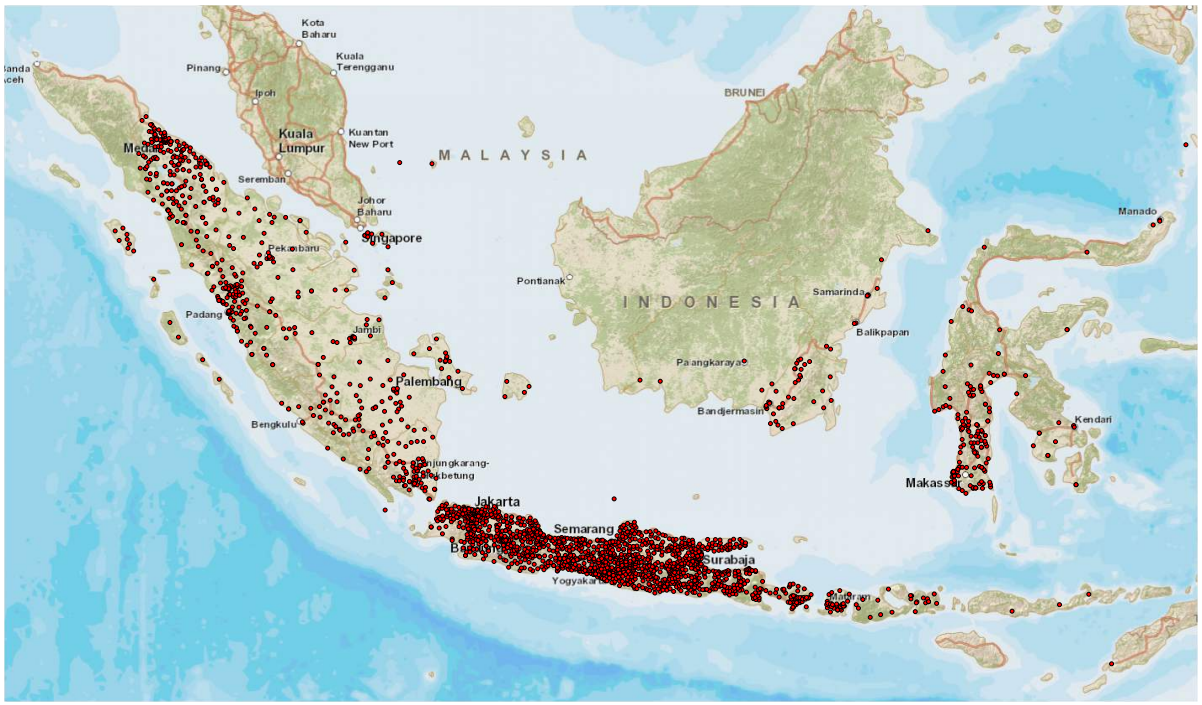
Panel

Crime data

- Crime and violence data: National Violence Monitoring Survey (Barron, Jaffrey and Varshney, 2014, Bazzi, Koehler-Derrick and Marx, 2019, Bazzi and Gudgeon, 2019)
- Records incidents from more than 2 million local newspapers articles
- We classify incidents into different categories based on detailed description
- Aggregated at the subdistrict-year level
- Crime rate per 100,000 inhabitants

Weather data

- Center for Climatic Research of the University of Delaware
- Extrapolated weather station data matched with household locations
- Grids of 0.5 by 0.5 degree $\approx$ 50 x 50 km



Individual locations (dots) with weather data they are mapped to (squares)

Weather data

- Center for Climatic Research of the University of Delaware
- Extrapolated weather station data matched with household locations
- Grids of 0.5 by 0.5 degree $\approx$ 50 x 50 km

Individual locations (dots) with weather data they are mapped to (squares)

Summary statistics

Variable	Mean	Std. Dev.
Share male	0.47	0.09
Age	35.28	4.05
Household size	4.48	0.95
Years of education	8.64	2.09
Individuals of migrant origin at destination (%)	35.5	25.8
Precipitation (mm per month)	161.76	57.58
Yearly crime rate per 100,000 people		
Total crime	14.63	15.37
Economic crime	4.21	5.36
Violent crime	10.41	11.18
People injured in incidents	9.94	12.95
People assassinated in incidents	1.48	1.78
People sexually assaulted	1.72	2.88
People kidnapped	0.08	0.34
Number of observations	7,171	
Number of locations	1,013	

Note: Sources: Indonesia Family Life Survey, National Violence Monitoring System and University of Delaware.

OLS results

	(1) Total crime	(2) Economic crime	(3) Violent crime	(4) Individuals injured	(5) Individuals killed	(6) Individuals sexually assaulted	(7) Individuals kidnapped
Percentage of migrants	0.075***	0.039***	0.037*	0.081***	0.003	0.001	0.001
	(0.028)	(0.011)	(0.021)	(0.029)	(0.003)	(0.007)	(0.001)
Destination rainfall lagged	0.012	0.007	0.006	0.014	-0.002	0.006*	0.000
	(0.013)	(0.005)	(0.009)	(0.012)	(0.002)	(0.003)	(0.001)
Mean dependent variable	14.63	4.21	10.41	9.94	1.48	1.72	0.08
Location FEs	Y	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y	Y
R-squared	0.644	0.611	0.616	0.600	0.329	0.288	0.188
Observations	7,171	7,171	7,171	7,171	7,171	7,171	7,171
No. of locations	1,013	1,013	1,013	1,013	1,013	1,013	1,013

Note: Dependent variables are yearly rates per 100,000 inhabitants and were created by dividing the total number of crimes, economically-motivated crimes, violent crimes and number of individuals inured, killed, sexually assaulted and kidnapped in the subdistrict each year (according to the data collected by the National Violence Monitoring System) by the population in that subdistrict according to the 2010 population census. Percentage of migrants is constructed as the ratio of individuals not born in the subdistrict to the total number of individuals living in that subdistrict in each year, times 100. Regressions are weighted by the number of individuals in each location in each year, and standard errors are clustered at the destination level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Second stage results

	(1) Percentage of migrants	(2) Total crime	(3) Economic crime	(4) Violent crime	(5) Individuals injured	(6) Individuals killed	(7) Individuals sexually assaulted	(8) Individuals kidnapped
Origin rainfall lagged	-0.075***							
	(0.018)							
Percentage of migrants		0.44	0.44***	0.03	0.56	0.02	0.04	0.01
		(0.450)	(0.165)	(0.336)	(0.435)	(0.059)	(0.105)	(0.014)
Destination rainfall lagged	0.066***	0.01	0.00	0.01	0.01	-0.00	0.00	-0.00
	(0.017)	(0.011)	(0.005)	(0.009)	(0.010)	(0.002)	(0.003)	(0.000)
Mean dependent variable	35.28	14.63	4.21	10.41	9.94	1.48	1.72	0.08
F-stat of excluded instrument	17.66							
Location FEs	Y	Y	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y	Y	Y
Observations	7,171	7,171	7,171	7,171	7,171	7,171	7,171	7,171
No. of locations	1,013	1,013	1,013	1,013	1,013	1,013	1,013	1,013

Note: Column 1 shows the results of regressing the percentage of the population in the subdistrict from migrant origin on rainfall levels in the subdistrict's catchment area from the previous calendar year according to Equation 2 and the subdistrict's own rainfall level in the previous calendar year. For columns 2 through 8, the dependent variables are identical to those reported in Table 2. Regressions are weighted by the number of individuals in each location in each year, and standard errors are clustered at the destination level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Additional First Stage Lags Economic Crimes Violent Crimes

Results based on survey data

- Could there be reporting bias in news paper coverage of crime?
  - If so, could this bias be correlated with the share of migrants at a destination
- Complementary analysis with household survey data from the National Socioeconomic Survey
  - Nationally representative household survey conducted by Indonesian Central Bureau of Statistics
  - 2,045,225 individuals from 318 districts
  - Rich dataset with questions on crime victim, gender, education, migrant status, etc
- Data on being a crime victim and reporting to police

Second stage results using survey data

	(1) Crime victim
Percentage of migrants	-0.002
	(0.006)
Destination rainfall lagged	-0.000
	(0.000)
Observations	3,342
Location FEs	Yes
Year Fes	Yes
First-stage F-stat	14.92
Variable mean	0.0159

Robustness checks

Second stage results using survey data

	(1) Crime victim	(2) Reported to police
Percentage of migrants	-0.002	0.305***
	(0.006)	(0.065)
Destination rainfall lagged	-0.000	0.000*
	(0.000)	(0.000)
Observations	3,342	1,602
Location FEs	Yes	Yes
Year Fes	Yes	Yes
First-stage F-stat	14.92	14.92
Variable mean	0.0159	0.193

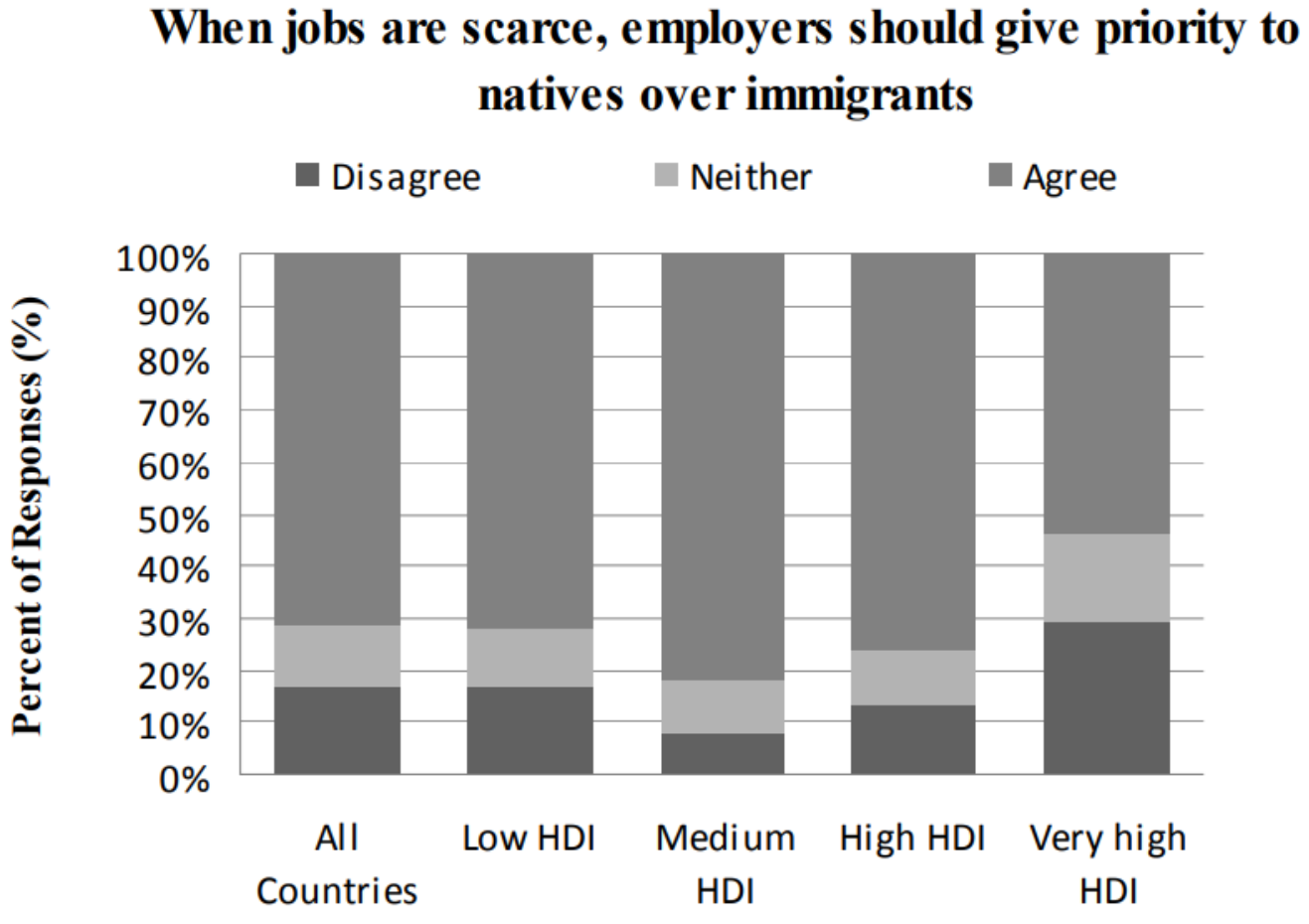
Robustness checks

Conclusion

- We estimate the causal effect on crime of internal migration in Indonesia using weather shocks in origin areas as instrument
  - 1 pp increase $\approx$ 6 percent increase in the share of migrants leads to a 3.9 percent increase in economic crimes
  - No effect on being a victim of crime
  - Increase in reporting the crime to the police, conditional on being a victim
- Importance of understanding data sources:
  - Could there be reporting bias in news paper coverage of crime?
  - Could this contribute to discrepancy between public opinion and empirical evidence?

Appendix

Public opinion about job preferences



Back

Possible effects of immigration on crime

- Theoretical predictions of effects of immigration on crime ex-anti ambigious (Spenkuch, 2014)
- Immigrants may be more likely to commit crime:
  - Immigration may have lower outside options (lower education/ wages/ employment)
  - Immigrants may belong to high-crime groups (e.g. young men)
- Immigrants may be less likely to commit crime:
  - May face deportation in addition to usual punishment
  - Immigrants may be positively selected, also in unobservable dimensions (enterpreneurship, motivation etc)
- Spillover effects
  - Immigration may worsen/improve natives' labor market conditions and hence affect their propensity to commit crime
- Cultural differences may lead to xenophobia and reduce integration (Card, Dustmann and Preston, 2012)

Back

Our context: Interal migration in Indonesia

- Theoretical predictions of effects of immigration on crime ex-anti ambigious (Spenkuch, 2014)
- Immigrants may be more likely to commit crime:
  - Immigration may have lower outside options (lower education/ wages/ employment)
  - Immigrants may belong to high-crime groups (e.g. young men)
- Immigrants may be less likely to commit crime:
  - May face deportation in addition to usual punishment
  - Immigrants may be positively selected, also in unobservable dimensions (enterpreneurship, motivation etc)
- Spillover effects
  - Immigration may worsen/improve natives' labor market conditions and hence affect their propensity to commit crime
- Cultural differences may lead to xenophobia and reduce integration (Card, Dustmann and Preston, 2012)

Back

First stage results

First stage relationship between origin precipitation and migrant share at destination

Origin rainfall current calendar year	-0.528***				-0.528***
	(0.172)				(0.173)
Origin rainfall previous calendar year		-0.750***		-0.745***	-0.773***
		(0.179)		(0.176)	(0.178)
Origin rainfall two calendar years ago			-0.195	-0.156	
			(0.151)	(0.150)	
Destination rainfall current calendar year	0.313*				0.283*
	(0.165)				(0.160)
Destination rainfall previous calendar year		0.656***		0.645***	0.667***
		(0.172)		(0.165)	(0.170)
Destination rainfall two years ago			0.19	0.152	
			(0.151)	(0.144)	
F-statistic of joint significance	9.38	17.66	1.67	11.92	9.4
Time fixed effects	Y	Y	Y	Y	Y
Destination fixed effects	Y	Y	Y	Y	Y
R squared	0.913	0.914	0.913	0.914	0.914
Observations	7,171	7,171	7,171	7,171	7,171
Number of locations	1,013	1,013	1,013	1,013	1,013

Back

Economic crimes

	(1) Robbery	(2) Plunder	(3) Theft	(4) Kidnap	(5) Drug Traffic	(6) Human Traffic
Percentage of migrants	0.23**	0.15***	0.06	0.02*	-0.00	-0.00
	(0.102)	(0.044)	(0.054)	(0.012)	(0.007)	(0.001)
Destination rainfall lagged	0.00	-0.00	-0.00	-0.00	-0.00	0.00*
	(0.003)	(0.001)	(0.002)	(0.000)	(0.000)	(0.000)
Mean dependent variable	2.19	0.60	1.25	0.07	0.04	0.00
Location FEs	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y
Observations	7,171	7,171	7,171	7,171	7,171	7,171
No. of locations	1,013	1,013	1,013	1,013	1,013	1,013

Note: Dependent variables are yearly rates per 100,000 inhabitants and were created by dividing the total number of each type of crime that occured in the subdistrict in every year (according to the data collected by the National Violence Monitoring System) by the population in that subdistrict according to the 2010 population census. Percentage of migrants is constructed as the ratio of individuals not born in the subdistrict to the total number of individuals living in that subdistrict in each year, times 100. Regressions include rainfall levels at the destination in the previous calendar year. Regressions are weighted by the number of individuals in each location in each year, and standard errors are clustered at the destination level.

Back

Violent crimes

	(1) Murder	(2) Arson	(3) Mistreat	(4) Molestation	(5) Fight	(6) Rape	(7) Destruction	(8) Shooting
Percentage of migrants	0.02	-0.03	-0.00	0.03	0.10	-0.01	-0.06*	0.02
	(0.036)	(0.022)	(0.143)	(0.032)	(0.095)	(0.087)	(0.037)	(0.045)
Destination rainfall lagged	-0.00	0.00	0.00	-0.00	-0.00	0.01	0.00	0.00
	(0.001)	(0.001)	(0.004)	(0.001)	(0.003)	(0.003)	(0.001)	(0.002)
Mean dependent variable	0.75	0.15	3.54	0.44	2.08	1.99	0.73	0.64
Location FEs	Y	Y	Y	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y	Y	Y	Y
Observations	7,171	7,171	7,171	7,171	7,171	7,171	7,171	7,171
No. of locations	1,013	1,013	1,013	1,013	1,013	1,013	1,013	1,013

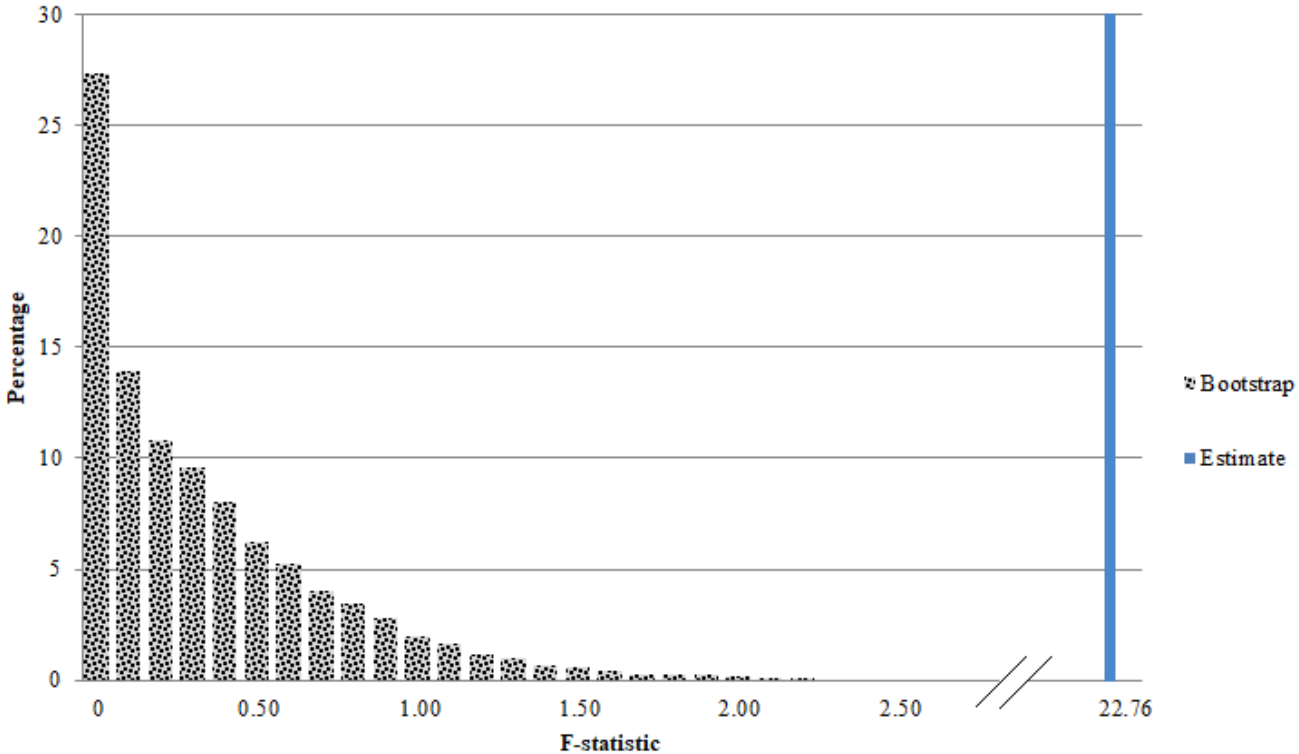
Note: Dependent variables are yearly rates per 100,000 inhabitants and were created by dividing the total number of each type of crime that occured in the subdistrict in every year (according to the data collected by the National Violence Monitoring System) by the population in that subdistrict according to the 2010 population census. Percentage of migrants is constructed as the ratio of individuals not born in the subdistrict to the total number of individuals living in that subdistrict in each year, times 100. Regressions include rainfall levels at the destination in the previous calendar year. Regressions are weighted by the number of individuals in each location in each year, and standard errors are clustered at the destination level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Back

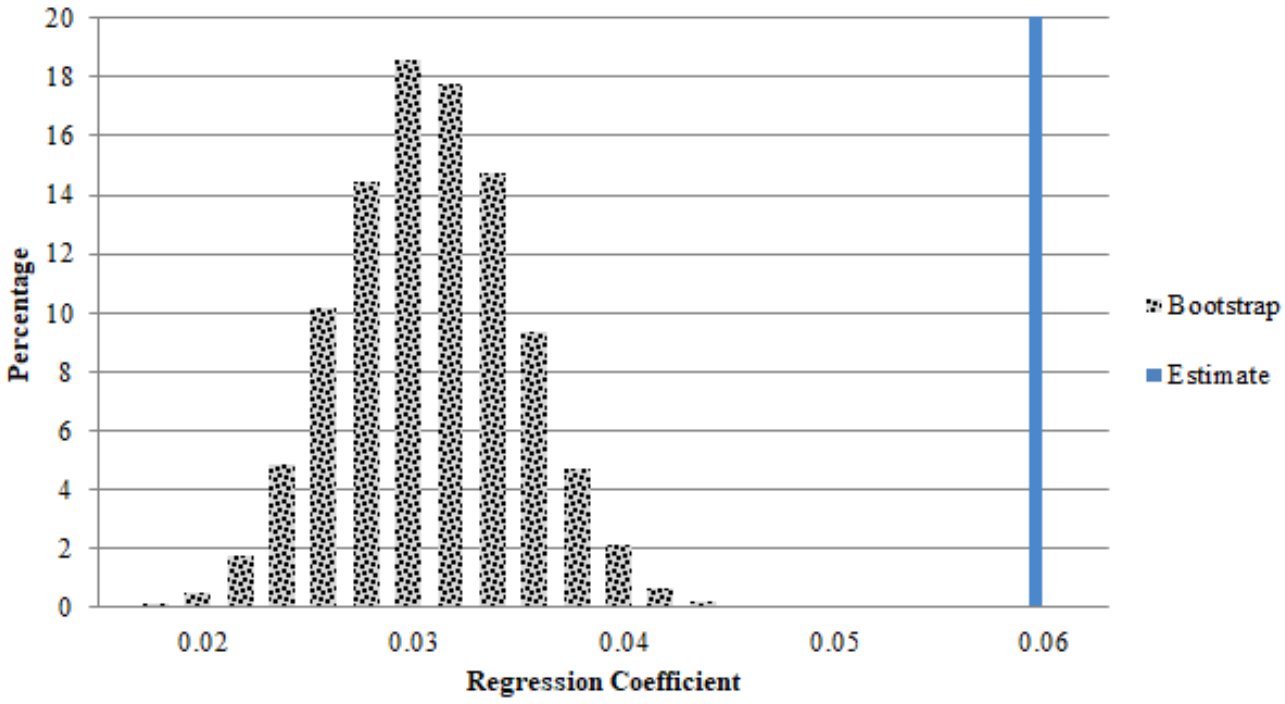
Testing the Exclusion Restriction

- If our exclusion restriction is invalid and trade patterns (or any other relationship) are not coincident with migration patterns, we expect different weighting schemes on catchment area rainfall to produce similar estimates
- We test this hypothesis by bootstrapping rainfall weights
  - For each destination, we fix the bootstrap catchment area to be the empirical migration catchment area we observe
  - Bootstrap the weighted origin rainfall measure by drawing a set of weights from a uniform distribution
- If it is the case that our migration rainfall measure is simply proxying for some correlated activity that takes place within the catchment area, we may expect many of these alternate weighting schemes to produce a similar relationship

Testing the Exclusion Restriction



Testing the Exclusion Restriction



Back

Long Distance Migration

- The exclusion restriction would be violated if rainfall at the origin affects the destination via channels other than migration
  - If the areas are economically connected through the movements of goods rather of people
  - Specifically, good rainfall conditions at the origin could increase the supply and affordability of agricultural trade into urban areas, which could stimulate eocnomic development.
- Tests this alternative channel by comparing our main results to those obtained when only considering migration that is at least 100 kilometres in distance
- If results were driven by local trade, there would be no reason to believe that these effects would be stronger over longer distances

Back