An application of structural equation model to analyse impacts of environmental factors on migration in Burkina Faso

Yajing Zhu [1] Florence de Longueville [2] Sabine Henry [3]

[1]Department of Statistics
London School of Economics and Political Science, UK

[2]Department of Geography
University of Liege, Belgium

[3]Department of Geography
University of Namur, Belgium

LSE Social Statistics Seminar, 9 March, 2017

Motivation

- Theory of climate-induced migration in rural areas.

  Concensus in the geography literature: environmental factors influence human migration (Koubi et al., 2016) , but little is known about the relative importance of climate factors and other factors (e.g. socio-economic drivers)
- Conceptual framework proposed by (Black et al., 2011)
- Burkina Faso (BK), West Africa.

  BK exposed to a diverse range of climates and has been previously identified as potentially vulnerable to climate-induced displacement.

Conceptual framework

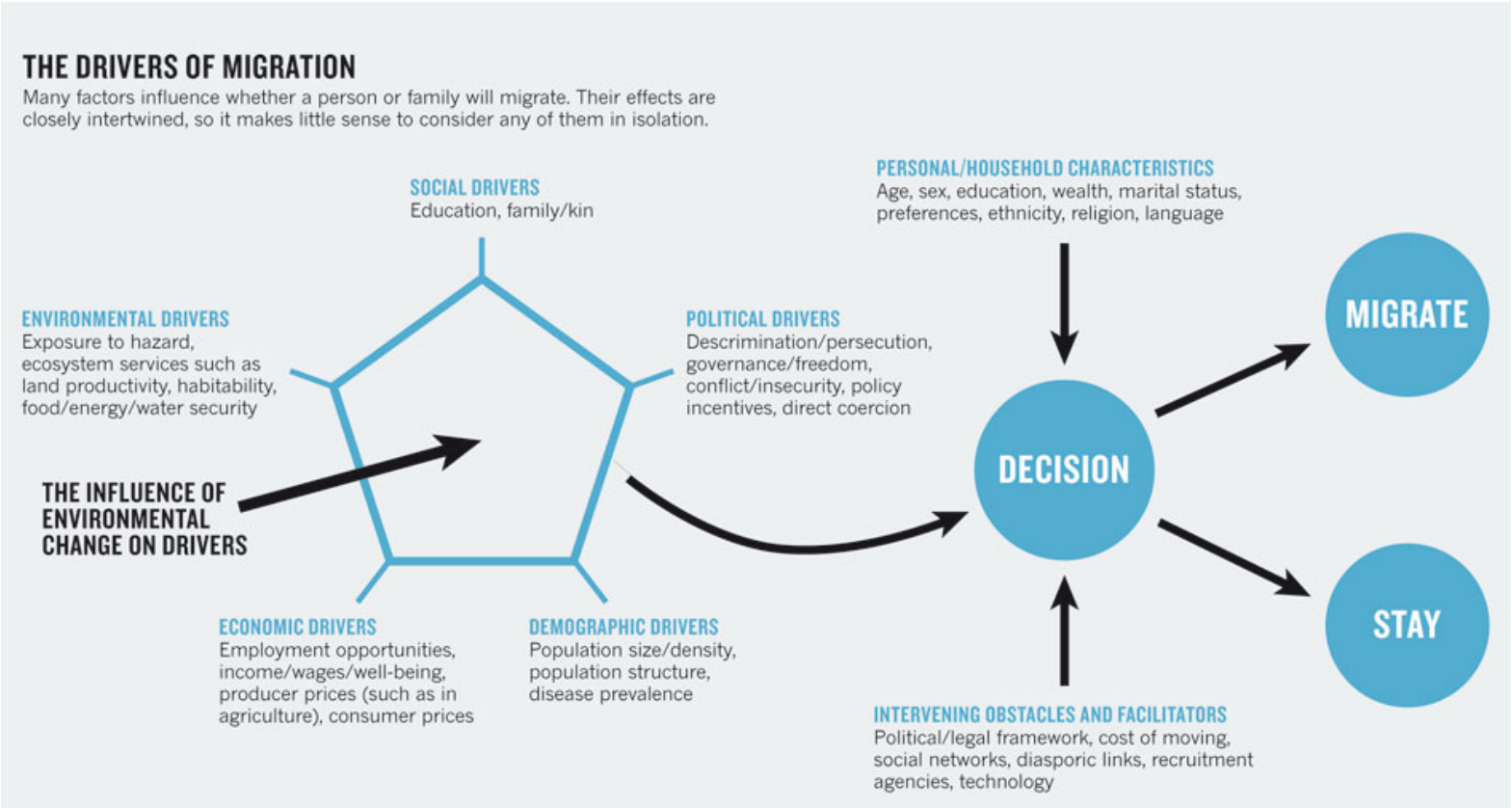


Figure: Conceptual framework showing the drivers of migration and the influence of environmental change on migration decision (souce: Black et al., 2011)

Research questions

Main interest

How do multiple drivers of migration decision (with a particular interest on environmental factors) interact with each other and jointly contribute to individual's decision to migrate?

1. Do adverse environmental conditions influence individual's decision to move (short-term or long-term)?
2. Are these effects of [1] mediated by socio-economic drivers?
3. Is the mechanism discovered in [1] and [2] different for male and female sub-populations?

Previous work

- Survival analysis: e.g. use multilevel event history analysis to estimate the impacts of rainfall conditions on the hazard of first migration (S. Henry, Schoumaker, and Beauchemin, 2004)

  Finding 1: temporary rainfall deficits (droughts) $\uparrow$ the risk of long-term migration to rural areas; $\downarrow$ the risk of short-term migration to more distant destinations (e.g. rural to urban)

  Finding 2: environmental factors do affect out-migration decisions, though the effects are both modest and more complex than anticipated (Bilsborrow and S. J. Henry, 2012).

- Multiple regressions with a common set of environmental or socio-economic predictors (the most commonly used approach by demographers interested in climate change)

  Common limitation: assuming perfect measurement by using one or two indicators; effect of environmental factors on SES is not often discussed

Our approach

Construct a structural equation model (Muthén, 1984) to model the association structure among latent and observed variables.

- Can reduce the dimension of various indicators by introducing latent summary variables $\Rightarrow$ focus on main quantities of interest.
- Can be generalised to more complex structures (e.g. multilevel, longitudinal, mixed-type data).
- Can explore/test other research questions.

Data I

- A retrospective survey with multiple modules; conducted in 2000-2001.
- Environmental data (linked to survey data using GIS): rainfall indicators: extracted from two climate databases.(1) Monthly time-step for 1901-2000 (interpolation); (2) Daily rainfall records from 70 weather stations spreading out the country.
- Community-level data (a separate survey; later linked to individual and household data)

  collected in 600/1,700 villages and towns. Topics covered: land availability, transportation, health centers, schools, employment opportunities, and agricultural mechanisms.

Data II

- Household-level data
  - massive missing data (blurred memory about household situation at first migration)
    - information on income level, family size, household amenities is not available

- Individual-level data
  - 8,644 men and women aged 15-64 yrs, from 800 locations
  - migration history from age 6 up to the interview time
  - records of each spell of residence $\geq$ 3 months
  - covered topics: employment, marital and birth status.

Working dataset

Working dataset: 4,394 individuals from 793 villages/towns.

- Migration window: first out-migration during 1980-2000 when the individual is $\geq$ 15 years old.
- Key response variables (S. Henry, Schoumaker, and Beauchemin, 2004):
  - Decision to migrate (binary)
  - Duration of stay after 1^{st} migration (3-category): duration is calculated as the time between the month of arrival at the 1^{st} village after 1^{st} migration and the month of arrival at the 2^{nd} village.
    - Non-migrants: No records of migration up to the interview time;
    - Short-term migrants: Duration $\in$ (3 months, 2 years);
    - Long-term migrants: Duration $\geq$ 2 years.

Note: censoring is not yet considered here; ST migrants could be LT migrants if they are followed.

Missing data

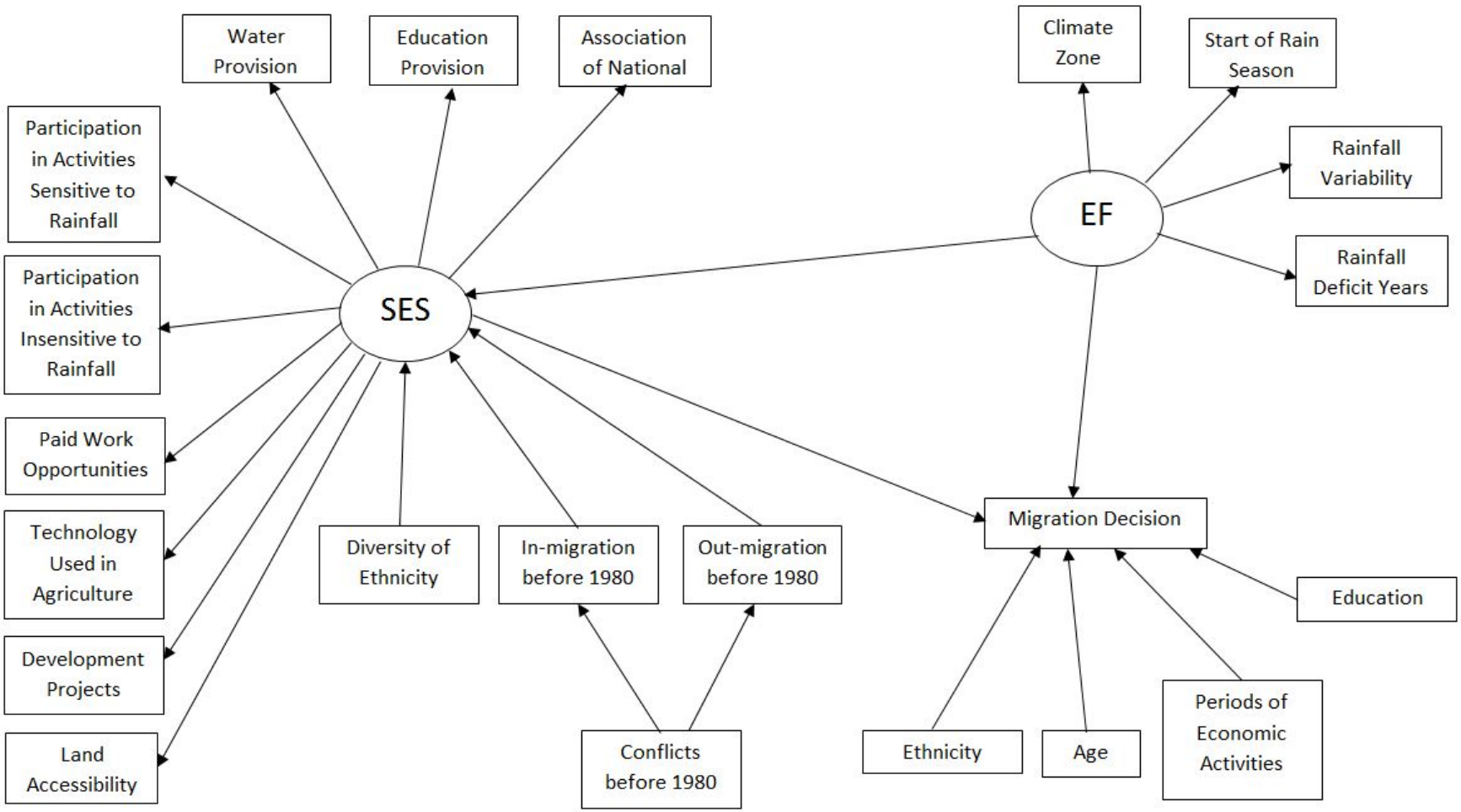
- Community-level: 12.2% of all sampled villages.
  - Random hot-deck imputation. Records missing at a particular village are replaced by records from a similar village (matching-sheet, based on density category and location).
- Individual-level: no missing data. The timing of employment, marital status and fertility status were double checked with administrative records by local researchers.

Summary of variables used in the SEM

Socioeconomic Factors	
Water Provision	Ordinal: 0,1,2 (2 =desirable)
Education Provision	Ordinal: 0,1,2 (2=desirable)
Association of National	Ordinal: 0,1,2 (2=good)
Land Accessibility	Ordinal: 0,1,2 (2=good)
Participation in Activities Sensitive to Rainfall	Ordinal: 0,1,2 (2=highly involved)
Participation in Activities insensitive to Rainfall	Ordinal: 0,1,2 (2= high degree of diversification)
Paid Work Opportunities	Ordinal: 0,1,2 (2=good)
Project Development	Ordinal: 0,1,2 (2=desirable)
Technology Used in Agriculture	Ordinal: 0,1,2 (2=advanced)
Political Factors	
Existence of Conflicts Before 1980	Binary: 0,1 (1=exist)
Demographic Factors	
Diversification of Ethnicity	Ordinal: 0,1,2 (2=good)
Existence of In-migration Before 1980	Binary: 0,1 (1= exist)
Existence of Out-migration Before 1980	Binary: 0,1 (1=exist)
Environmental Factors	
Annual Rainfall	Ordinal: 1,2,3,4,5 (5=highest total annual rainfall)
Start of the Rain Season	Binary: 0,1 (1=not started)
Variability of Rainfall	Ordinal: 1,2,3,4,5 (5=highly variable)
3-year Deficit of Rainfall	Ordinal: 0,1,2,3 (3=highest)
Personal Characteristics	
Ethnicity	Binary: 0,1 (1=Mossi; 0=other groups)
Age	Continuous: log(age at 1st migration)
Number of Periods in Economic Activities	Ordinal: 1,2,3 (3=good)
Education Level	Ordinal:0,1,2,3 (3= highest)
Gender	Categorical: 1,2 (1=male)
Decision to Migrate	Dependent Variable (in the 1st analysis) Binary: 0,1 (1=migrate)
Duration of Stay	Dependent Variable (in the 2nd analysis) Nominal: 0,1,2 (0=non-migrants; 1=3 months to 2 years; 2=above 2 years)

SEM structure

SES and EF are community-level variables for the origin village.



Model specification

Using matrices

- Structural model (Key interest)

$$g(P_k^M) = \beta_{0k} + \boldsymbol{\beta}_{1k}^T \mathbf{X}^M + \beta_{2k}\eta^{SES} + \beta_{3k}\xi^{EF}$$
$$\eta^{SES} = \mathbf{B_0} X^{SES} + \mathbf{B_1}\xi^{EF} + \zeta$$

- Measurement model (the underlying variable specification)

$$\mathbf{X}^{*SES} = \boldsymbol{\alpha}_X^{SES} + \mathbf{A}_X \eta^{SES} + \boldsymbol{\delta}^*$$
$$\mathbf{Y}^{*EF} = \boldsymbol{\alpha}_Y^{EF} + \mathbf{A}_Y \xi^{EF} + \boldsymbol{\epsilon}^*$$

  - P_k^M=P(Duration of stay=k)
  - g=link function, multinomial logit
  - allow for clustered (community-level) standard errrors

Key results

Report results of main interest (after standardising w.r.t variances of latent/underlying latent variables) estimated in Mplus 6.1 - specify MODEL INDIRECT, MONTE CARLO re-sampling approach for SE(indirect effect).

Pathway	Estimates					
	Male		Female		Whole Population	
Duration of Stay	Short term	Long term	Short term	Long term	Short term	Long term
EF→						
Annual Rainfall	-1.77**		-1.41**		-1.59**	
Rainfall Variability	2.31**		1.86**		2.02**	
Number of Rainfall Deficit Years	1.06**		0.92**		0.97**	
Start of Rain Season (constrained)	/		/		/	
SES←EF	-0.61**		-0.21**		-0.41**	
Duration of Stay (versus "no migration"; in odd ratios)←						
EF	1.48	1.12	1.38**	1.41	1.44**	1.31
SES	0.89	1.09	0.34**	0.72	0.13**	0.94
Age(in logarithm)	0.22**	0.11**	0.15**	0.04**	0.17**	0.06**
Ethnicity	1.43	3.25**	1.51**	2.14**	1.45**	2.49**
Periods of Economic Activities	1.55**	4.32**	1.73**	3.40**	1.66**	4.05**
Education Level	1.13	1.31**	1.16	1.00	1.12	1.08

Figure: Results for the parts of SEM corresponding to the main research interest. ** $p < 0.05$, * $p < 0.1$

More details I

Interpretation of L.V.

- High values of SES: origin village has good use of agricultural technology, good association of national, poor education provision and low degree of participation in rainfall-insensitive activities.
- High values of EF: poor environmental conditions.

Effects of EF on migration (direct effect):

- Poor environmental situations $\Rightarrow \uparrow$ ST migration and women are particularly sensitive to this, compared to men.

More details II

Effects of SES on migration:

- High levels of SES $\Rightarrow \downarrow$ ST migration

Effects of EF on migration through SES (indirect effect):

- Disadvantaged environmental situation $\Rightarrow \downarrow$ community's SES (people move to engage more in rainfall-insensitive activities and use less agricultural technology) $\Rightarrow \uparrow$ ST migration.

Effects of personal characteristics on migration:

- Elder individuals are less prone to move (either ST or LT); individuals active in economic activities are more likely to move (particularly LT migration).

Sensitivity analysis

How stable our results are in relation to the selection of environmental variables?

- Focus on potential changes in direct and indirect effects of *Environmental Factor* on *Duration of Stay*. This analysis is using the whole sample.

Environmental Variables (In the Model)	Direct Effect (OR)		Indirect Effect (OR)		Total Effect (OR)		Model Fit Index	
	Short term	Long term	Short term	Long term	Short term	Long term	AIC	BIC
All 4 variables	1.44**	1.31	2.31**	1.16	2.32**	1.28	56914	57317
Neglecting (Holding other variables untouched)								
"Start of Rain Season"	1.43*	1.25	2.38**	0.99	2.59**	1.24	133552	134006
"Number of Deficit Years"	1.65*	1.02	2.24*	1.06	2.81*	1.08	128491	128931
"Annual Rainfall"	1.17*	2.51	2.33*	0.93	2.22*	2.32	123583	124024
"Rainfall Variability"	Model Does Not Converge							

Figure: Sensitivity analysis: change of estimated direct and indirect effects of Environmental Factor on Duration of Stay due to the selection of environmental indicators. Effects are presented in odds ratios (OR). Model fit indices AIC and BIC are listed in the last two columns.

Some troubles

- No absolute model fit indices in ML (outcome is categorical) (no observed covariance matrix to compare the model predictions to)

  Changing the estimator to MV-WLS (continuous underlying variables) could help.
- Interpretation is tricky: 1 SD increase in EF $\Rightarrow \alpha$ SD decrease in SES; 1 SD increase in SES $\Rightarrow$ change in odds of ST or LT migration decision (paths are not comparable!).

Challenges

- Only considered push effects (environmental and origin village); partial effects are tested (pull effects from destination areas are not considered) - needs more collaborative work with demographers.
- Outcome variable is duration of stay in the destination village - not accounting for the distance from origin village. A bi-variate response could be considered?
- Getting back to descriptive statistics, we find that over 60% of males leaving the village after 15 years return within 10 years while the return rate of women is just 15%. How to account for both processes (any suggestions on potential models)?
- Geographers are also interested in modelling more out-migration episodes - difficult to identify households (multiple partners; over 3 partners). Any experience of handling such data?

References

- Bilsborrow, Richard E and Sabine JF Henry (2012). "The use of survey data to study migration–environment relationships in developing countries: alternative approaches to data collection". In: *Population and environment* 34.1, pp. 113–141.
- Black, Richard et al. (2011). "The effect of environmental change on human migration". In: *Global Environmental Change* 21, S3–S11.
- Henry, Sabine, Bruno Schoumaker, and Cris Beauchemin (2004). "The impact of rainfall on the first out-migration: A multi-level event-history analysis in Burkina Faso". In: *Population and Environment* 25.5, pp. 423–460.
- Koubi, Vally et al. (2016). "The role of environmental perceptions in migration decision-making: evidence from both migrants and non-migrants in five developing countries". In: *Population and Environment* 38.2, pp. 134–163.
- Muthén, Bengt (1984). "A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators". In: *Psychometrika* 49.1, pp. 115–132.