Discovering Diverse Paths, Linking Fragmented Ideas

An Empirical Approach to Integrating Migration Theories

Filiz Garip
Harvard University

“Every person is a different world....
They think differently.”

- Pedro, a former U.S. migrant in Mexico

There are 12 million Mexican-born in the United States, about half of them are undocumented.

Who are these migrants? What brings them here?

There is no single answer to these questions.

There is no single answer to these questions.

Consider the changes in the Mexico-US stream.

	1970s	1990s	2000s
men	.75	.65	.73
from central-west states	.67	.47	.25
to CA, TX, or IL	.87	.75	.55

What are the origins of these shifts?

What determines who migrates from Mexico to the United States?

individual desires to maximize income

family strategies to diversify risks to income

social ties to former or current migrants

MACRO		*MICRO*
wage differences b/w Mexico and the US	Neoclassical	individual desires to maximize income
uncertainty in the Mexican economy	New economics	family strategies to diversify risks to income
past migration b/w Mexico and the US	Cumulative causation	social ties to former or current migrants

MACRO Segmented labor markets

persistent demand for
migrant labor in the US

World systems

capitalist expansion from
the US to Mexico

MACRO Segmented labor markets *MICRO*

persistent demand for migrant labor in the US

?

World systems

capitalist expansion from the US to Mexico

?

What determines who migrates from Mexico to the United States?

Three key ideas:

- neoclassical model
- new economics
- cumulative causation

What determines who migrates from Mexico to the United States?

Three key ideas:

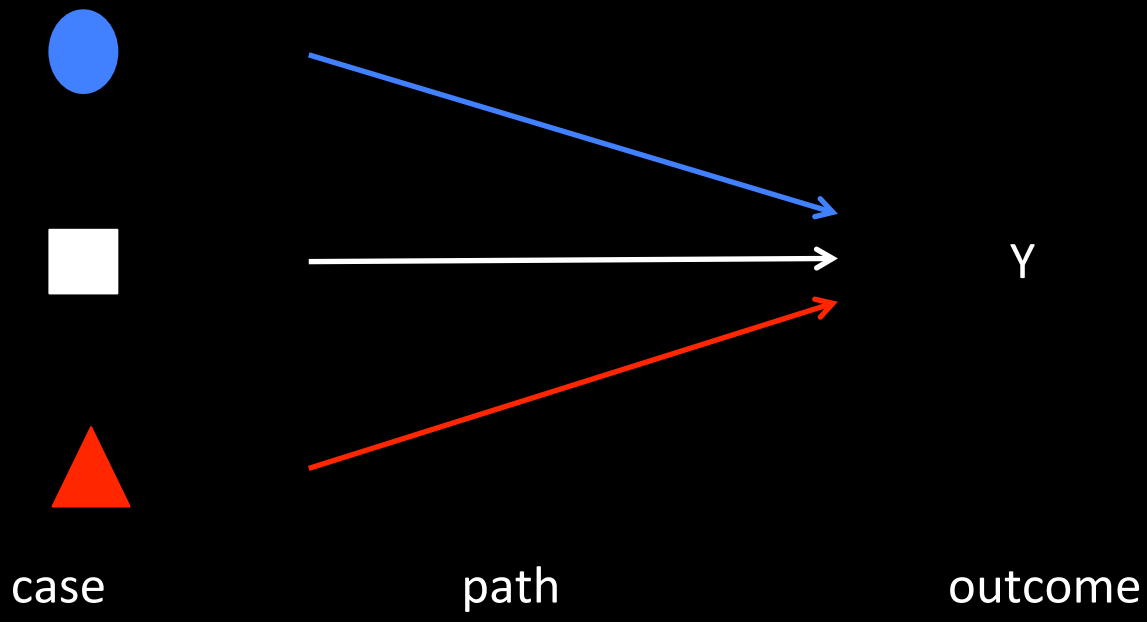
- neoclassical model
- new economics
- cumulative causation

When and for whom is each idea most relevant?

Argument

Migrants might have different reasons for coming to the United States.

case path outcome


Y
case
path
outcome

Most work focuses on the average migrant.

Most work focuses on the average migrant.

Regression analysis is the dominant method.

Most work focuses on the average migrant.

Regression analysis is the dominant method.

We mold questions/theories to fit the method.

Migration theories are conditional statements that apply in a specific context and to a specific subset of the population.

How do we discover the different paths
that bring Mexicans to the United States?

Strategy

1. Fix the outcome and study only the migrants.

Strategy

1. Fix the outcome and study only the migrants.

2. Search for different groups among migrants. Groups are defined by shared configurations of attributes. Assumption: Individuals with similar attributes face similar opportunity structures.

Strategy

1. Fix the outcome and study only the migrants.

2. Search for different groups among migrants. Groups are defined by shared configurations of attributes. Assumption: Individuals with similar attributes face similar opportunity structures.

3. Study the conditions that set apart each group from the other groups as well as non-migrants.

Who migrates?

When?

Why?

Data

Mexican Migration Project (MMP) surveyed about 200 randomly-selected households in 143 Mexican communities from 24 states between 1982 and 2013.

Our sample contains 19,243 migrants observed during their first U.S. trip between 1965 and 2010.

Method: Cluster analysis

Discovers groups with similar attributes in data

Method: Cluster analysis

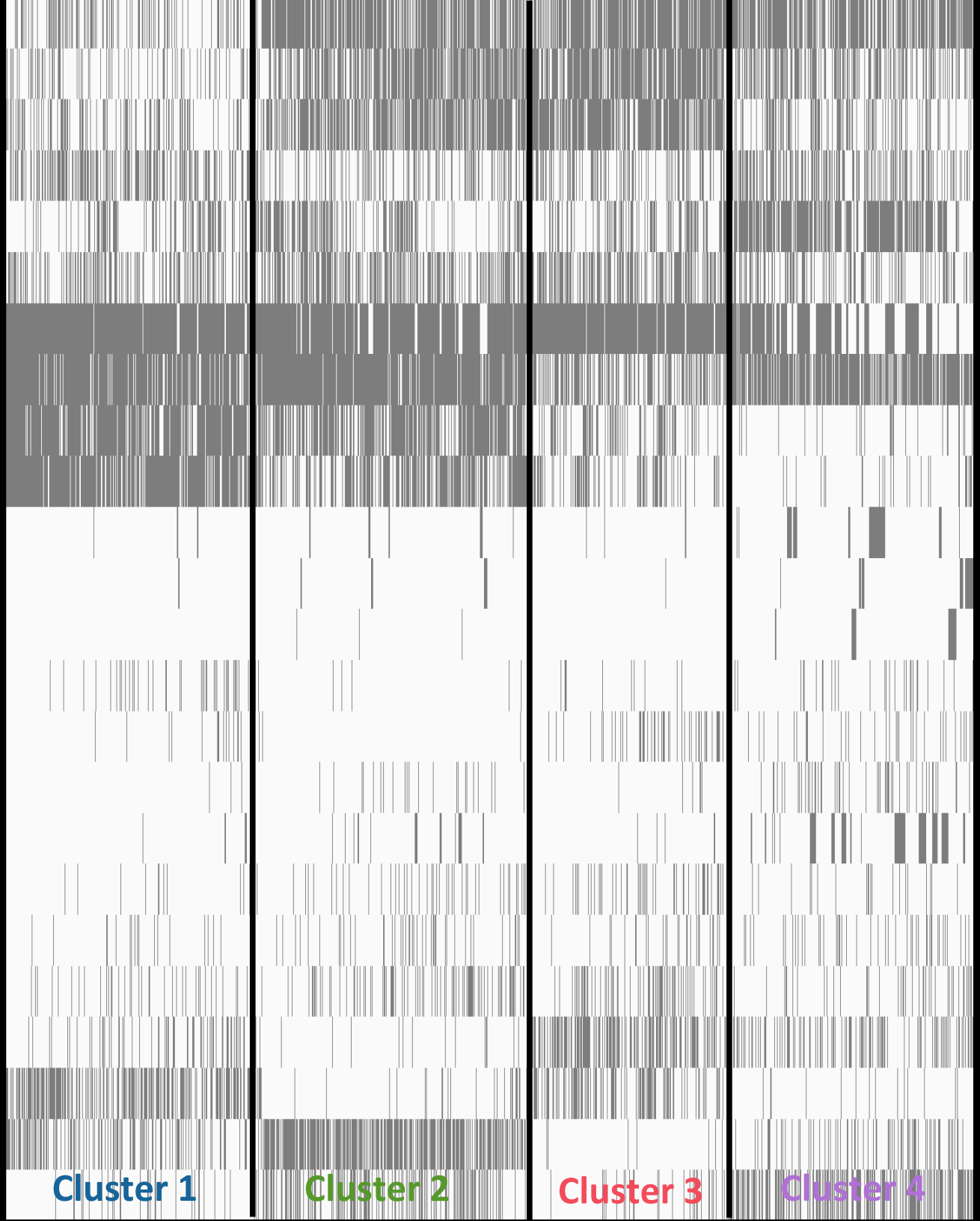
How it works:

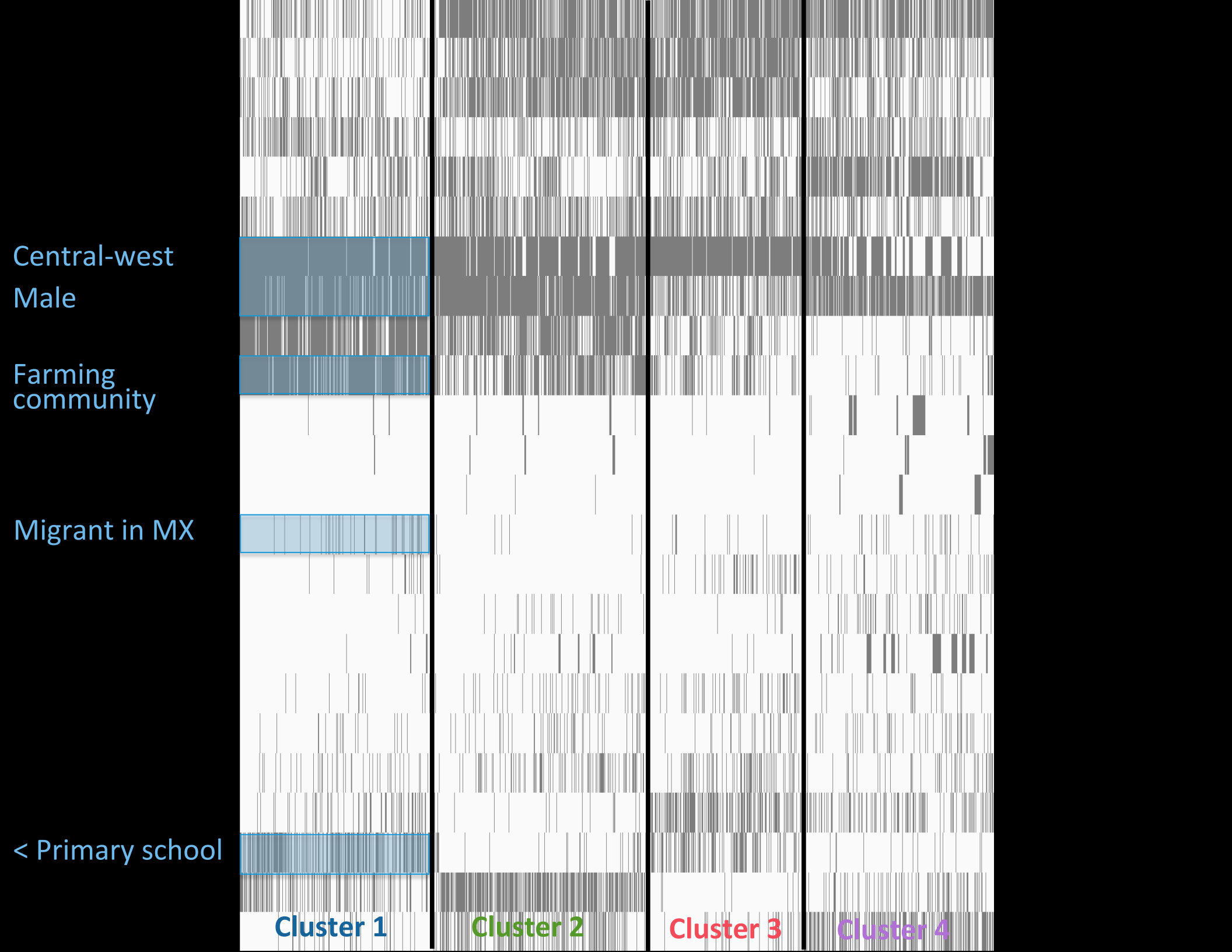
1. Choose and scale the relevant attributes
2. Choose an algorithm: K-means
3. Choose a similarity measure: City-block distance
4. Determine K, the number of clusters, using cluster validation measures

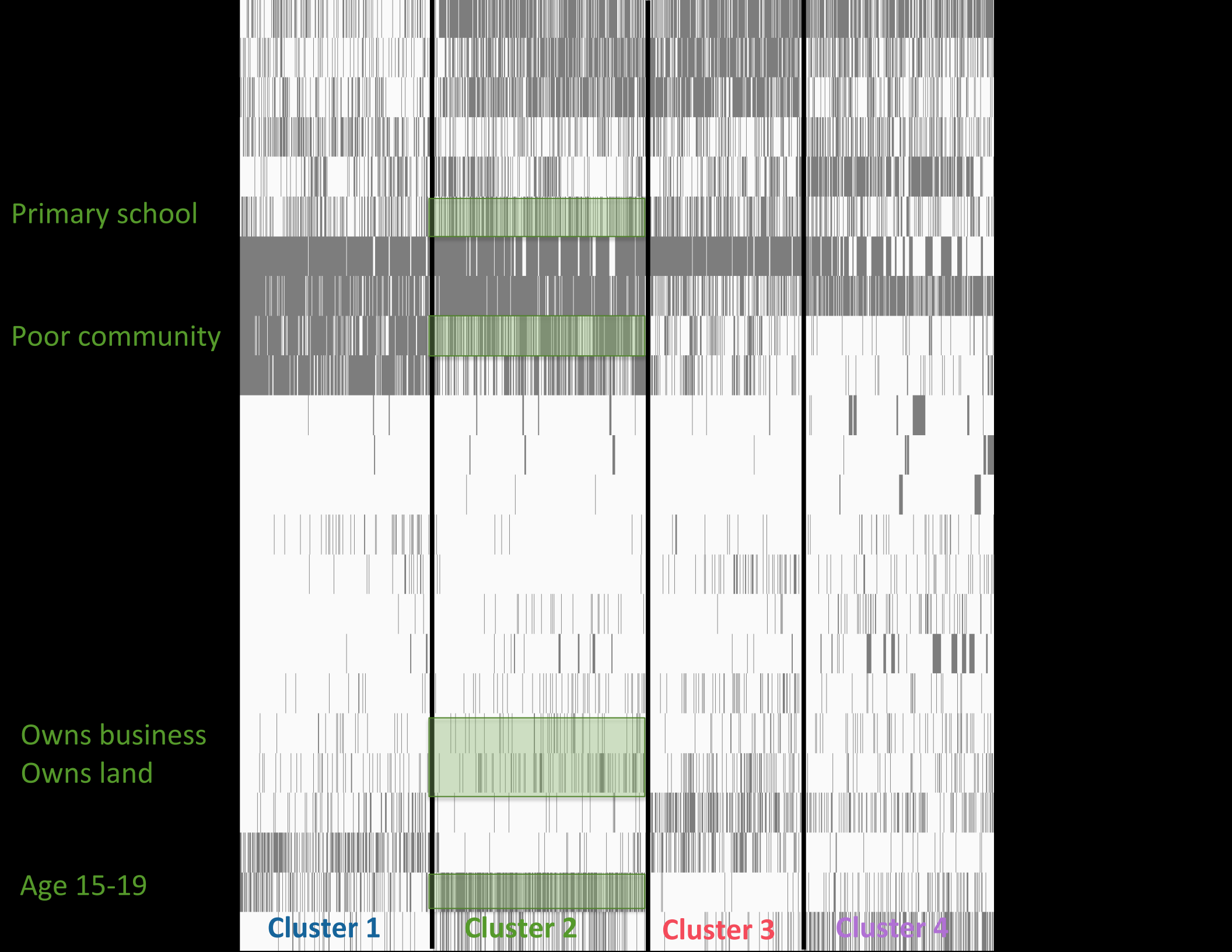
Who migrates?

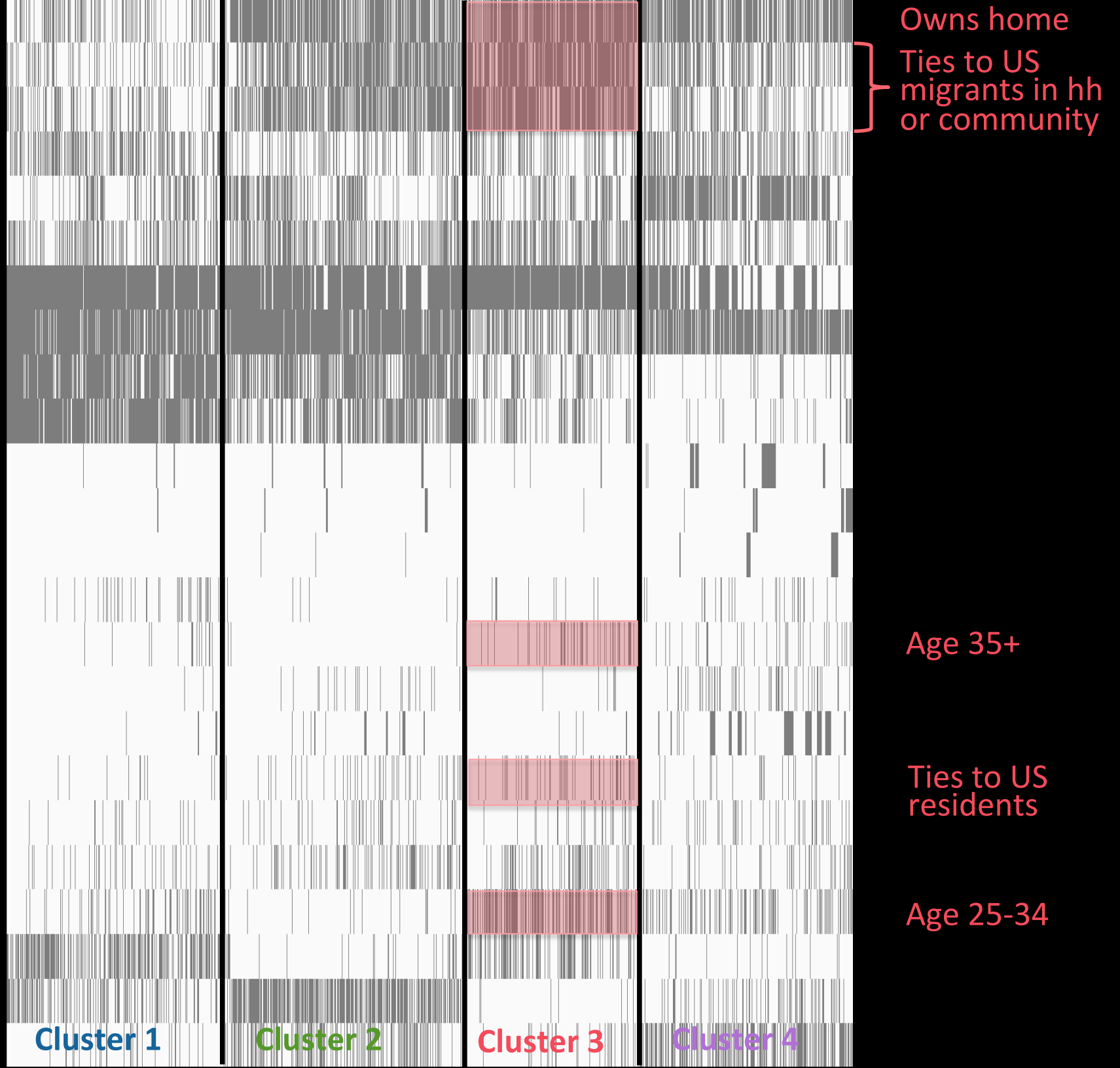
When?

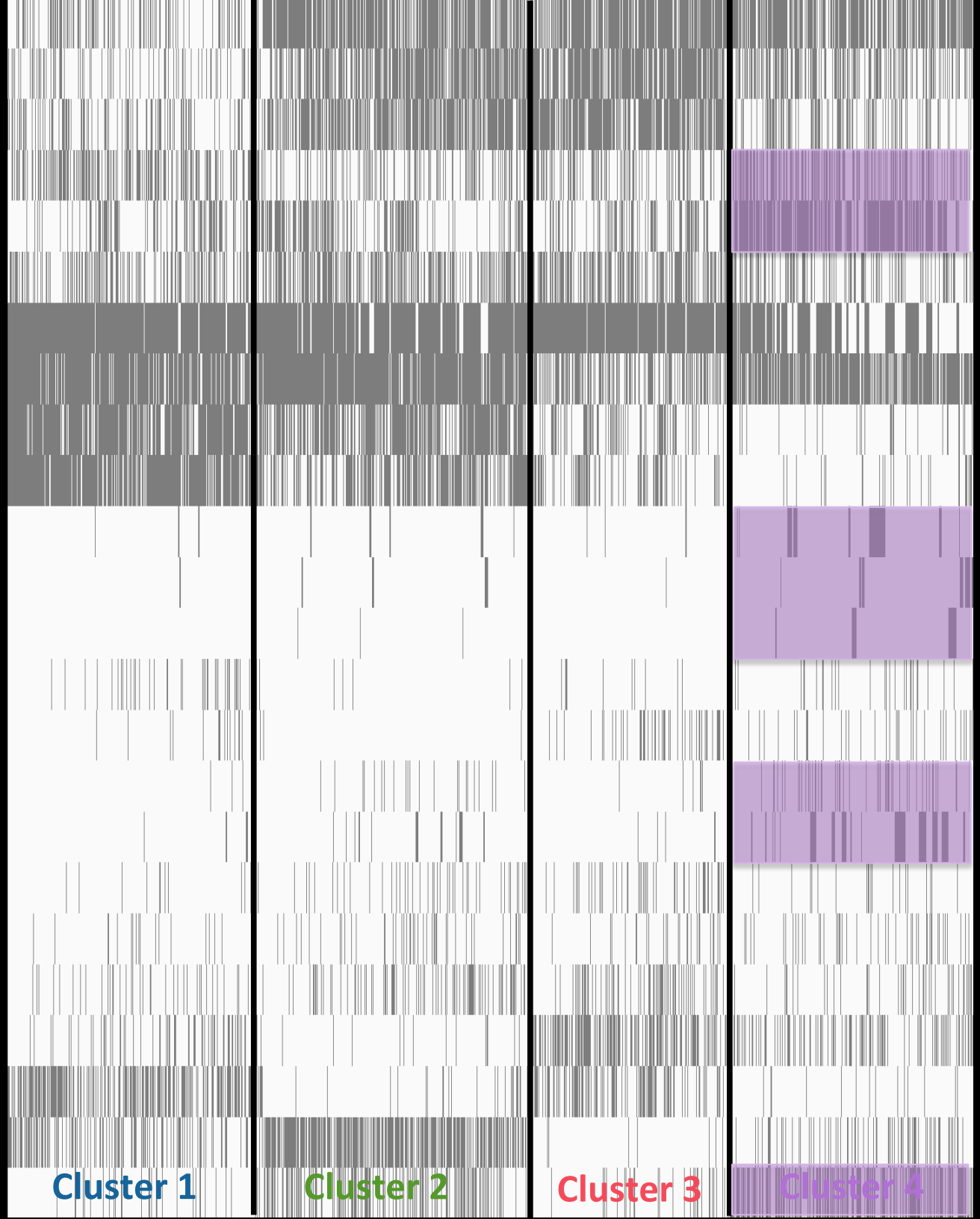
Why?


Cluster 1
Cluster 2
Cluster 3
Cluster 4

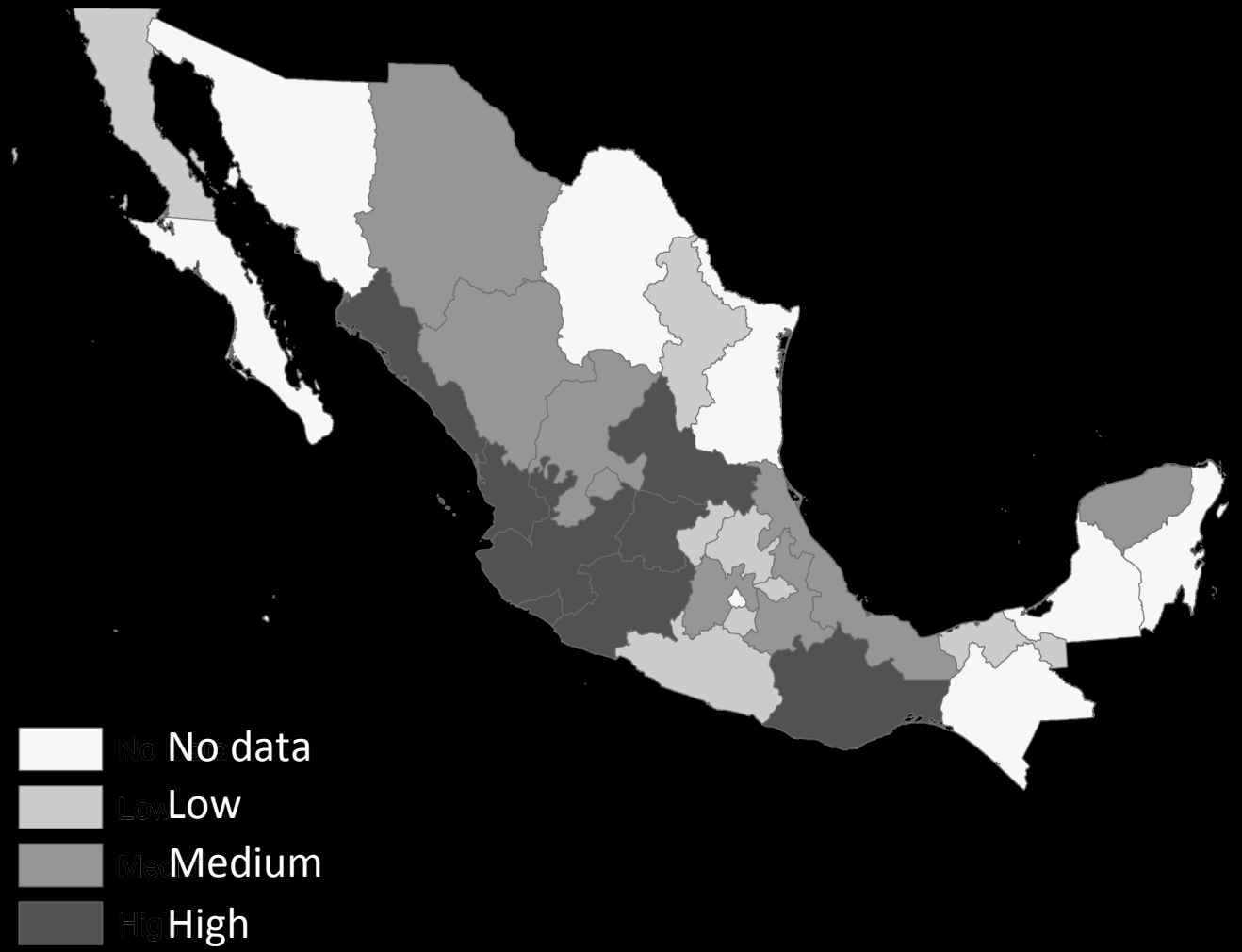

Central-west
Male
Farming community
Migrant in MX
< Primary school
Cluster 1
Cluster 2
Cluster 3
Cluster 4


Primary school
Poor community
Owns business
Owns land
Age 15-19
Cluster 1
Cluster 2
Cluster 3
Cluster 4

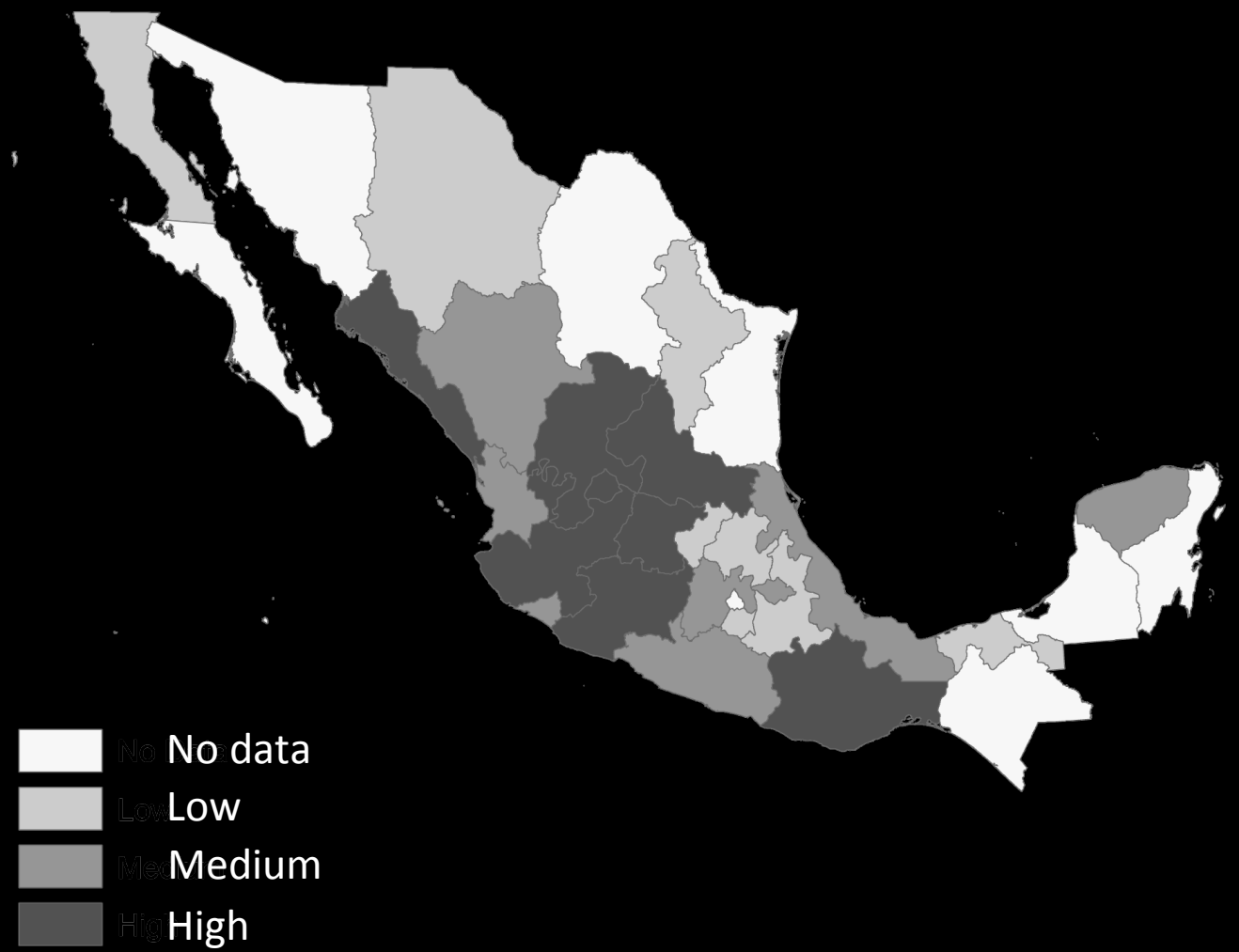

Owns home
Ties to US migrants in hh or community
Age 35+
Ties to US residents
Age 25-34
Cluster 1
Cluster 2
Cluster 3
Cluster 4


Age 20-24
Urban area
Border
Southeast
Central
High school+
Central-south
Middle school
Cluster 1
Cluster 2
Cluster 3
Cluster 4

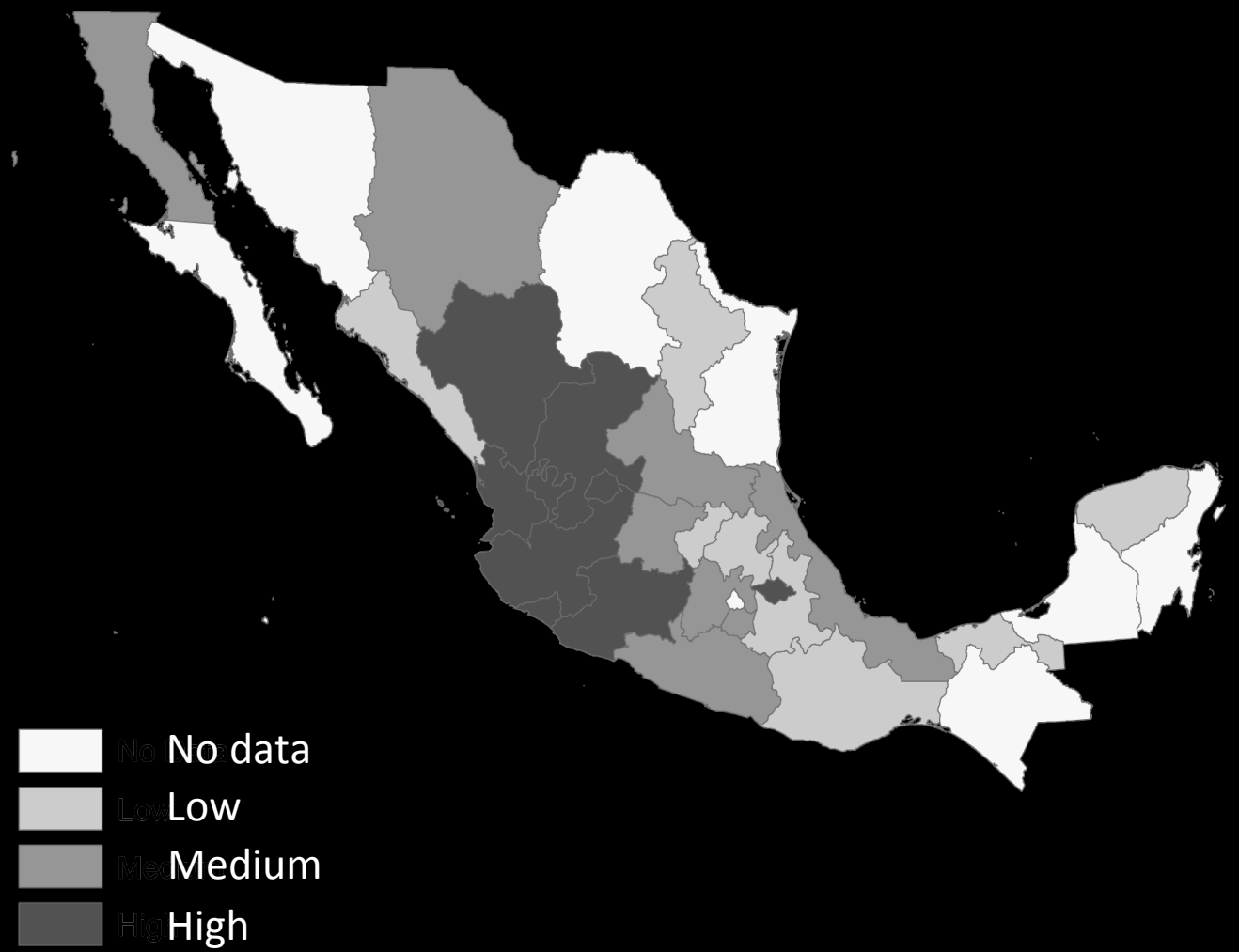
Cluster 1



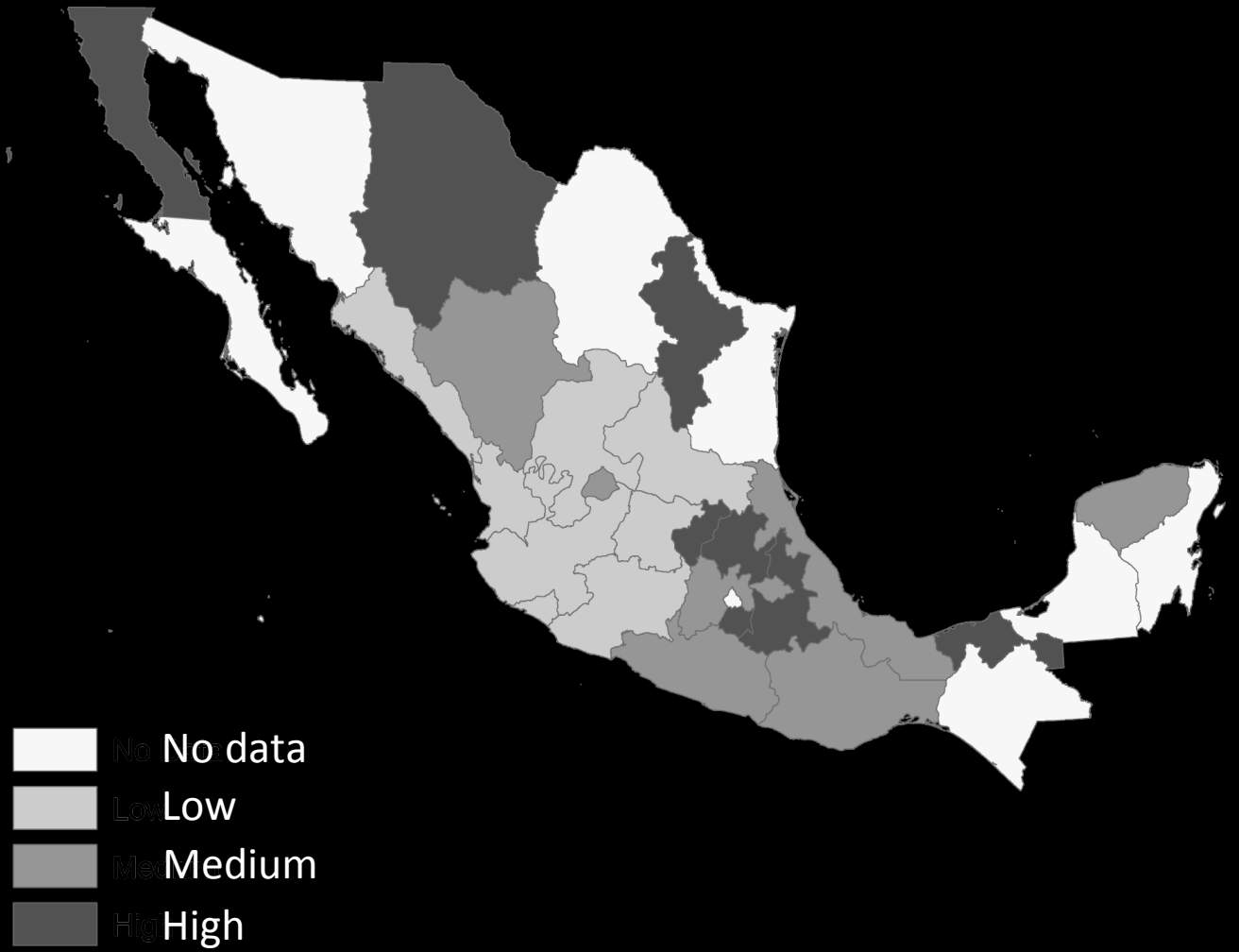
Cluster 2



Cluster 3



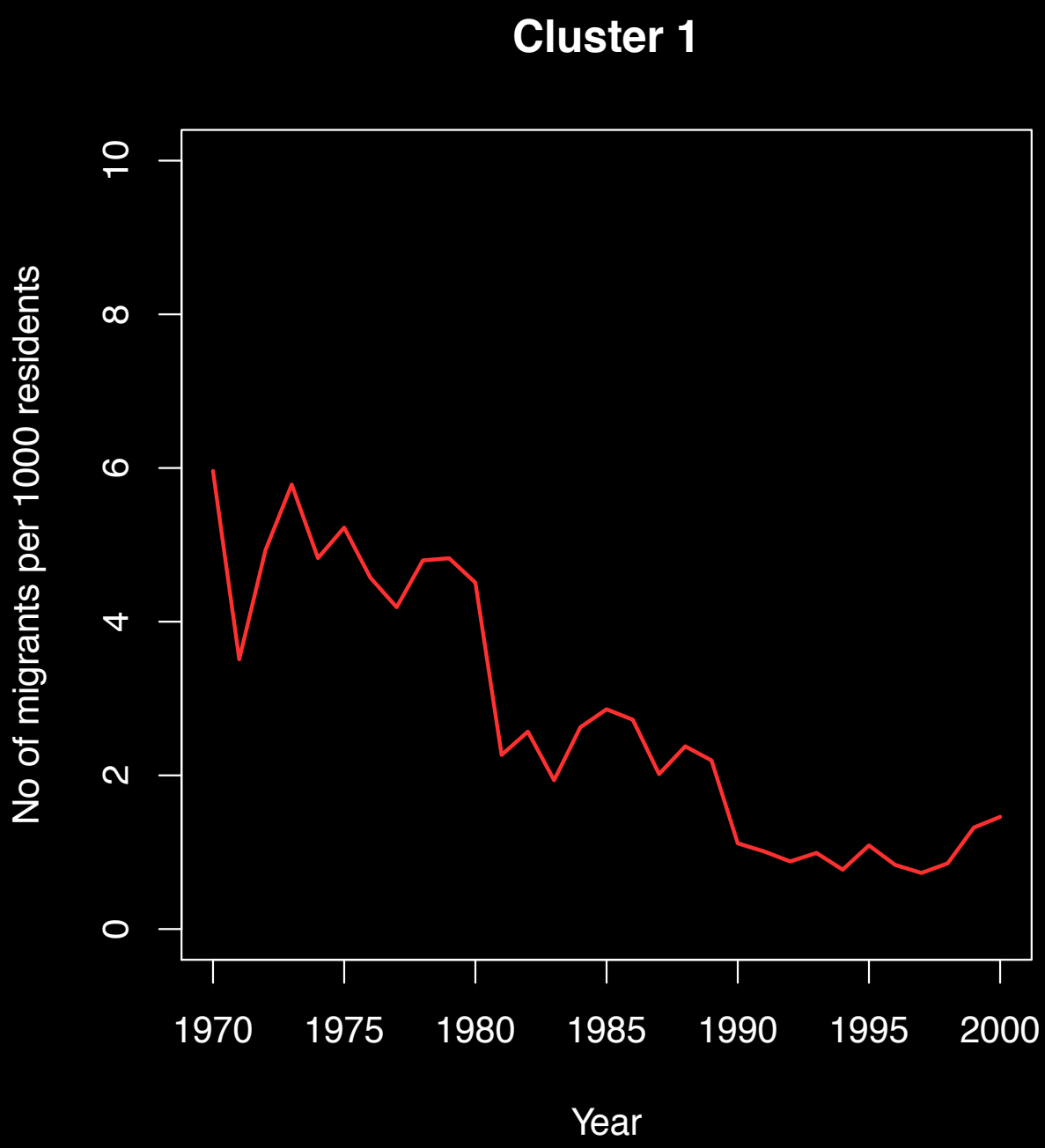
Cluster 4

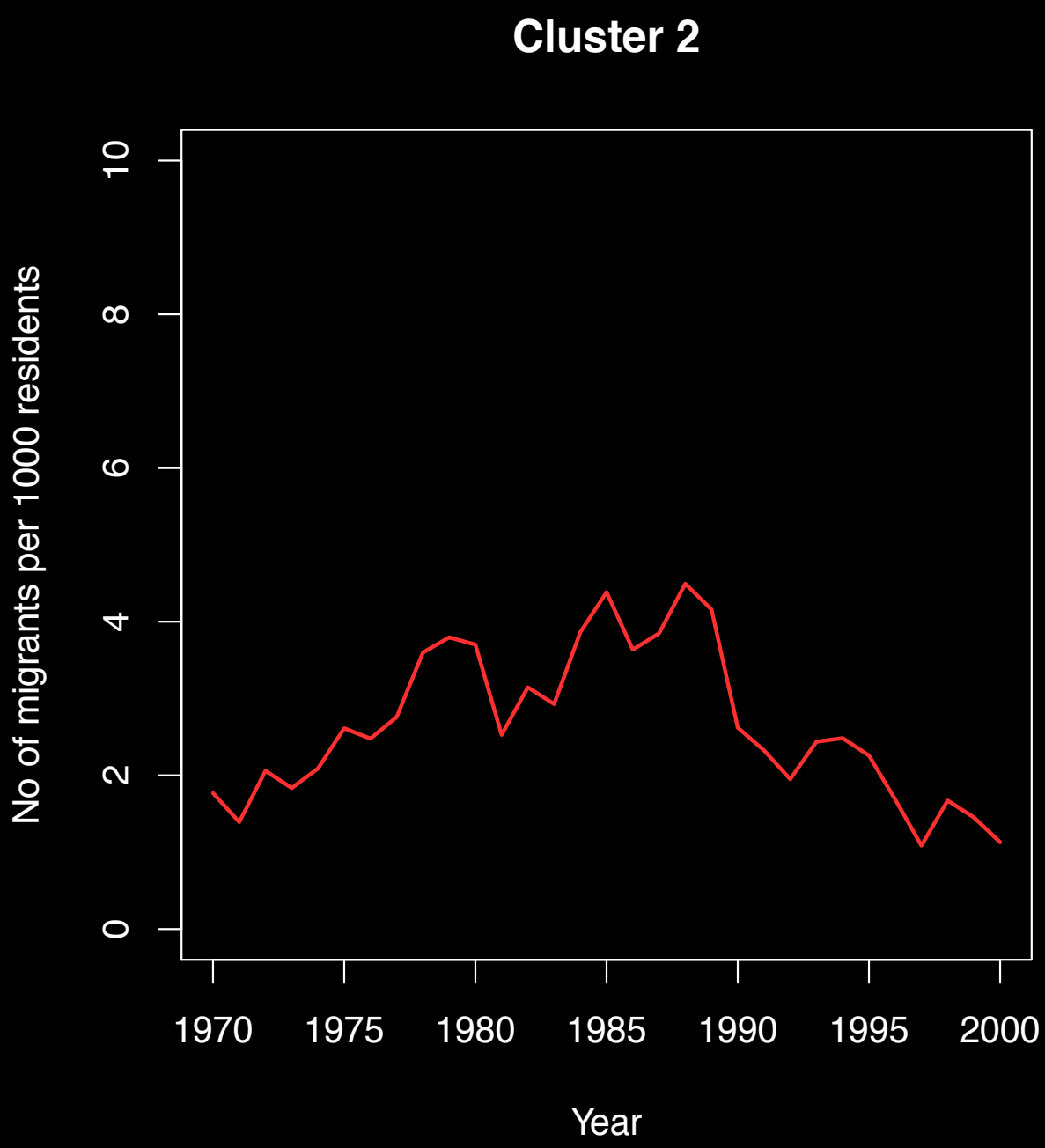


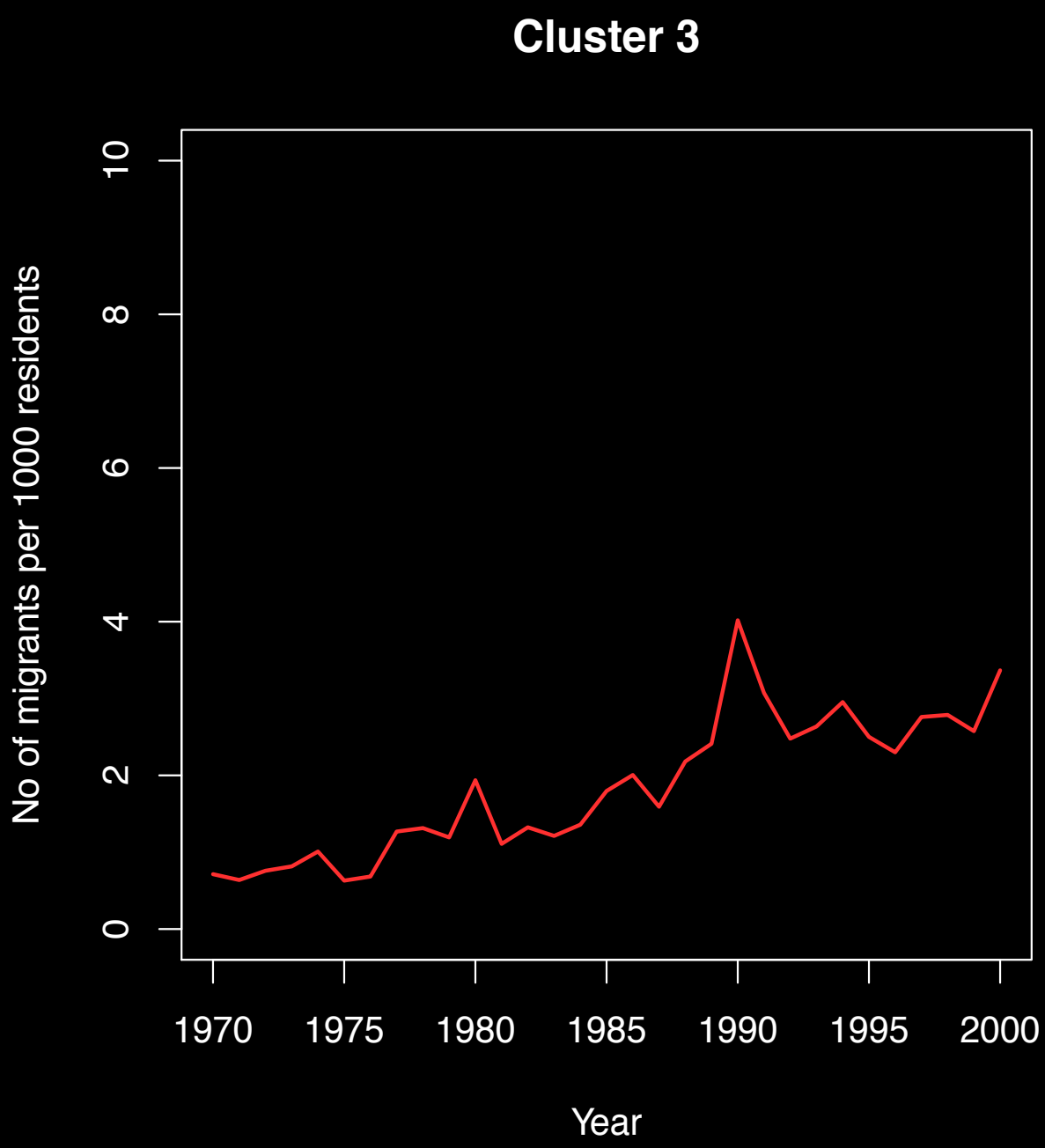
Who migrates?

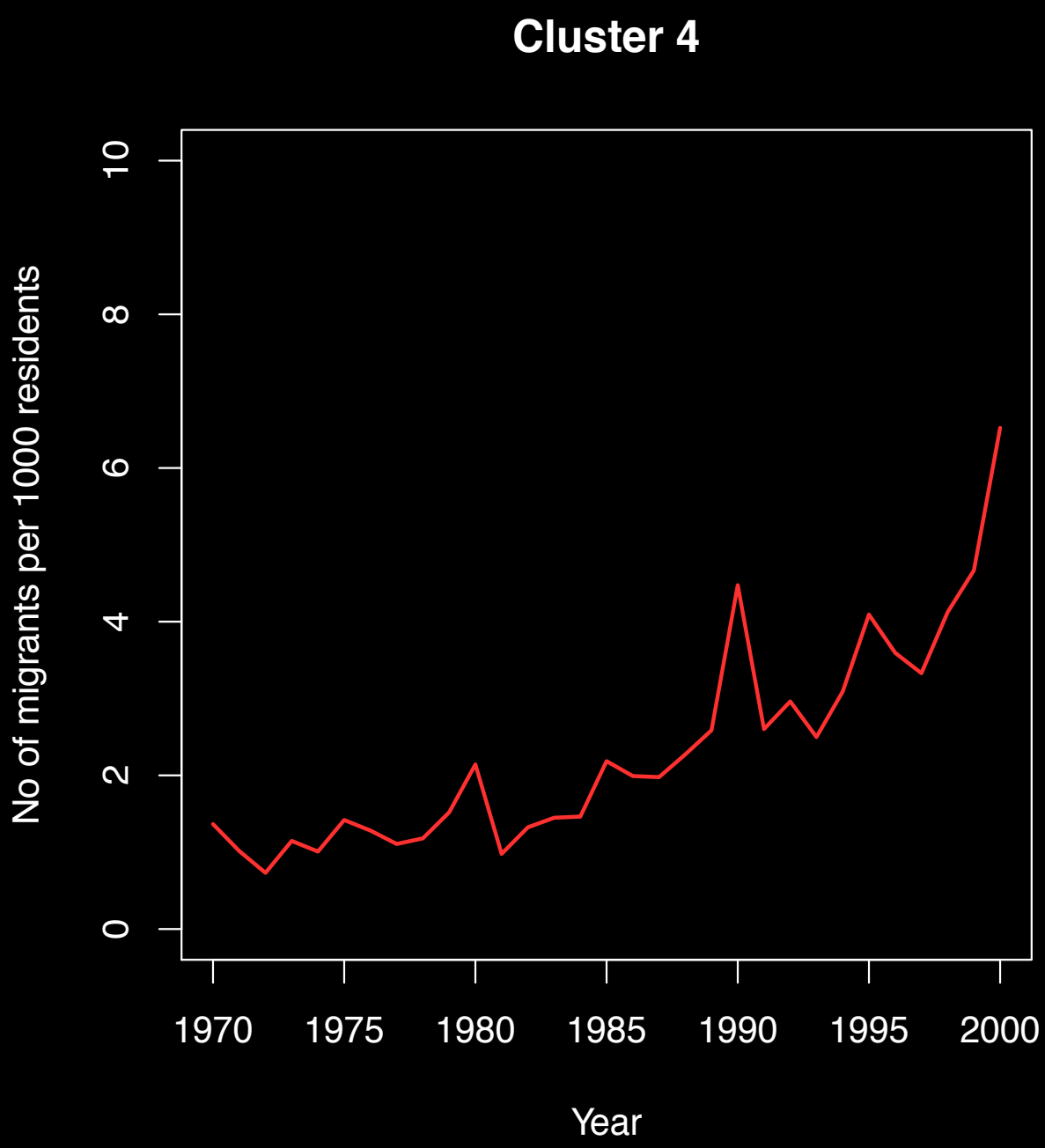
When?

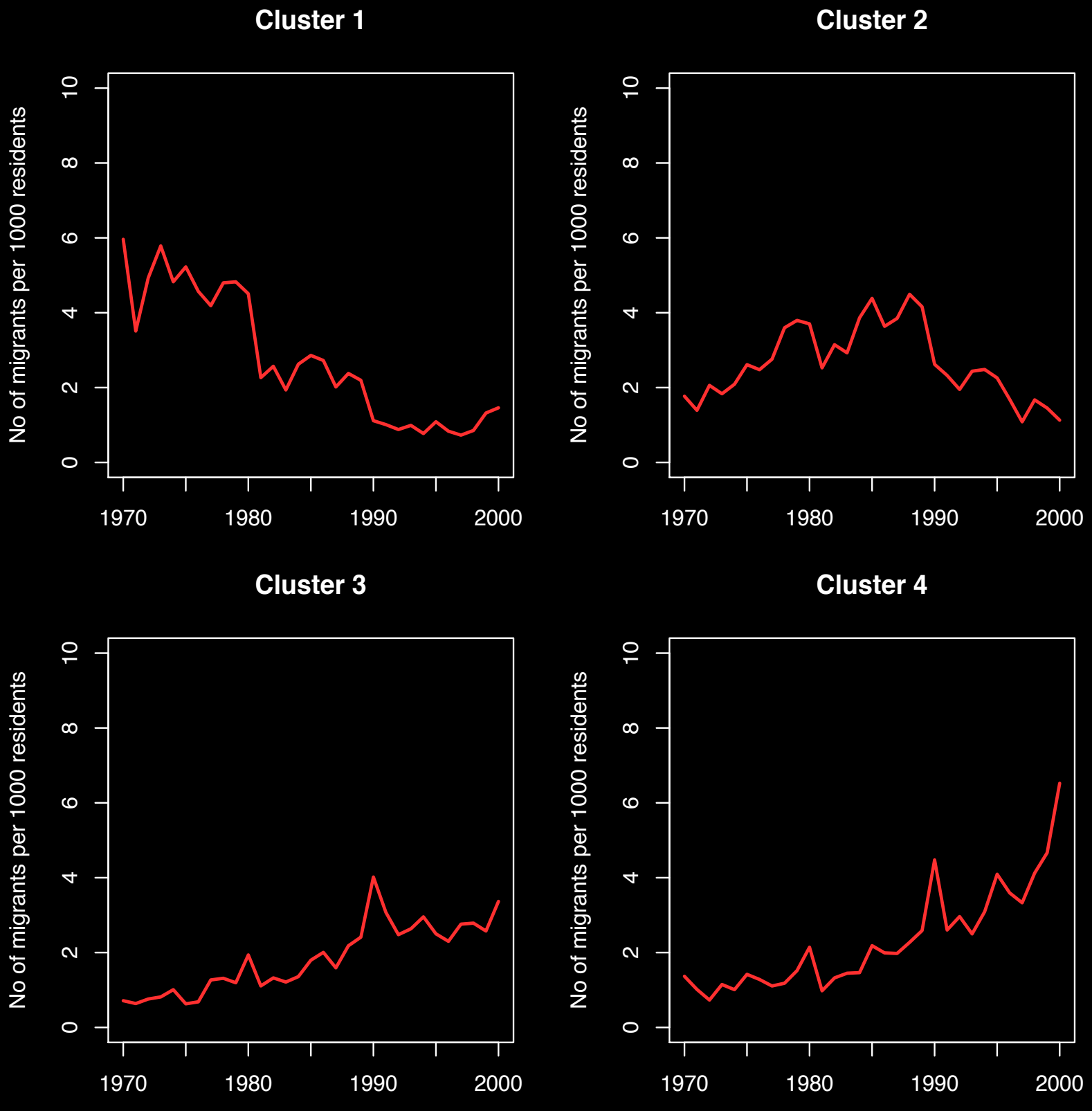
Why?


Cluster 1
No of migrants per 1000 residents
0
2
4
6
8
10
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 2
No of migrants per 1000 residents
0
2
4
6
8
10
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 3
No of migrants per 1000 residents
0
2
4
6
8
10
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 4
No of migrants per 1000 residents
0
2
4
6
8
10
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 1
Cluster 2
Cluster 3
Cluster 4
No of migrants per 1000 residents
0
2
4
6
8
10
1970
1980
1990
2000

Who migrates?

When?

Why?

Are different clusters responding to different macro-level conditions?

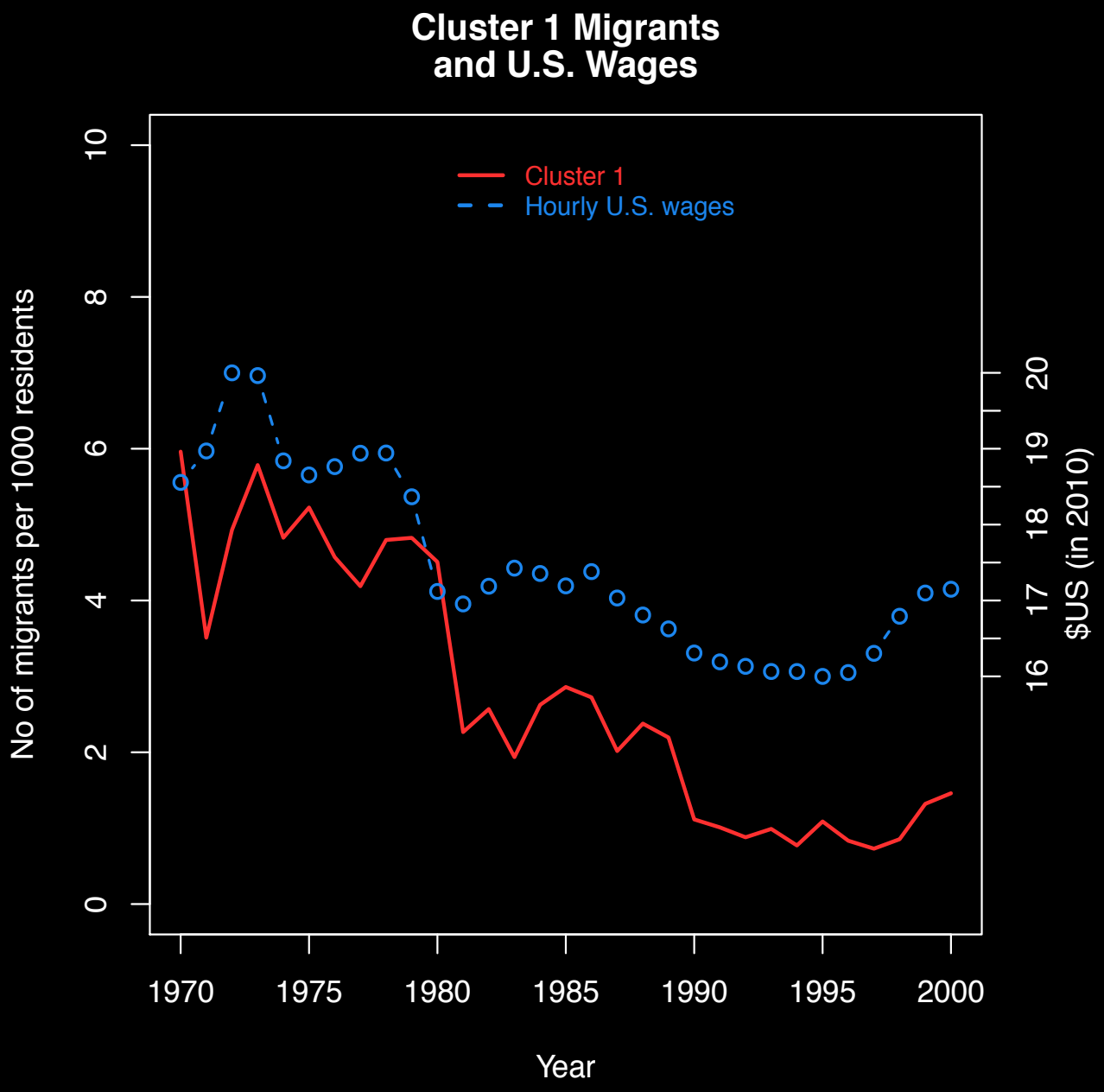
Neoclassical model	Wages in the US Wages in MX Unemployment in the US Unemployment in MX Border patrol enforcement (BPE) budget
New economics	Inflation in MX Δ in $ value of peso
Cumulative causation	Visa availability / family reunification
Segmented markets	Δ in employment in migrant-heavy sectors in US
World systems	MX-US trade

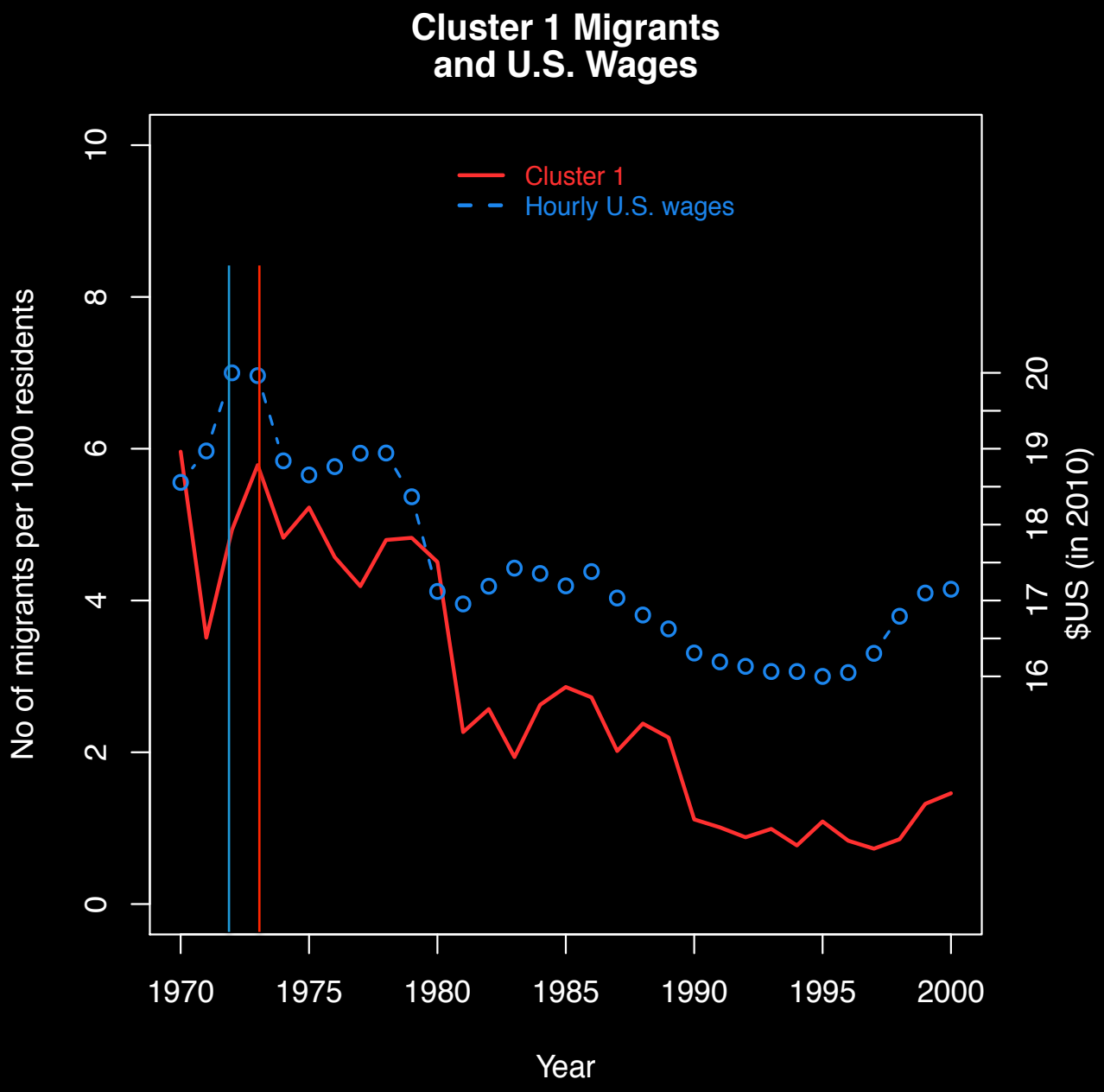
Are different clusters responding to different macro-level conditions?

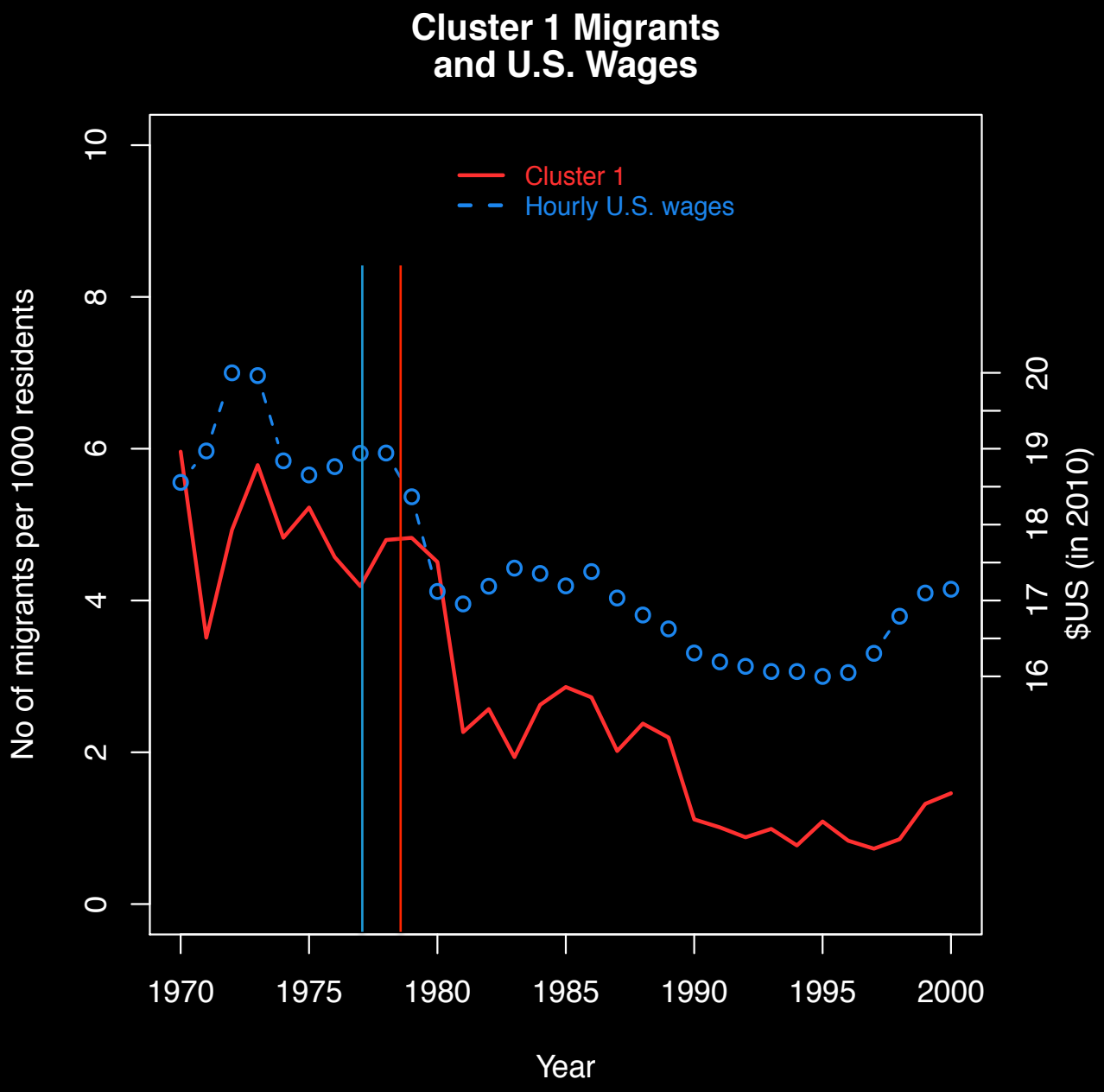
Cluster 1

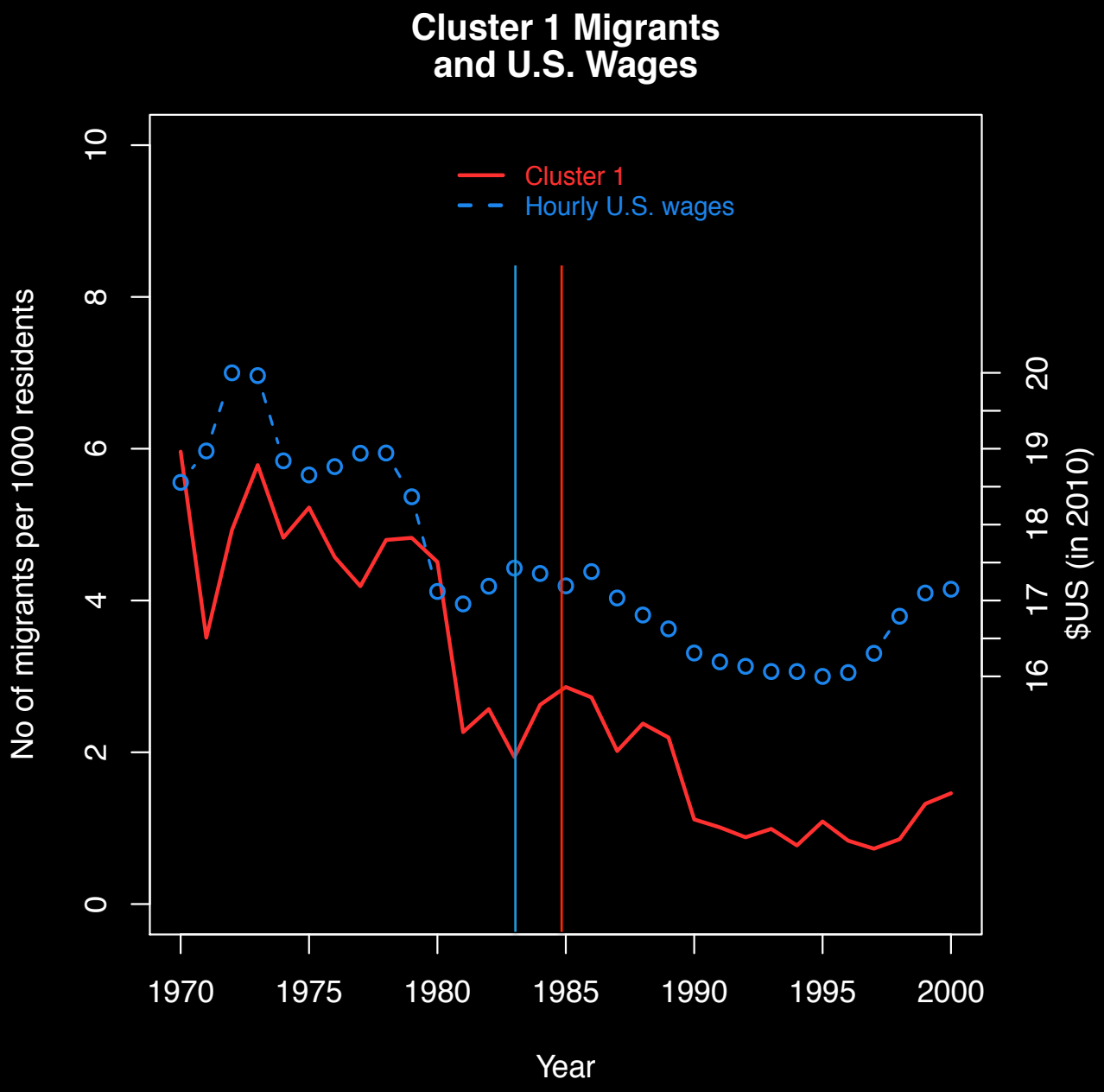
US low-skill wage	+
BPE budget	-

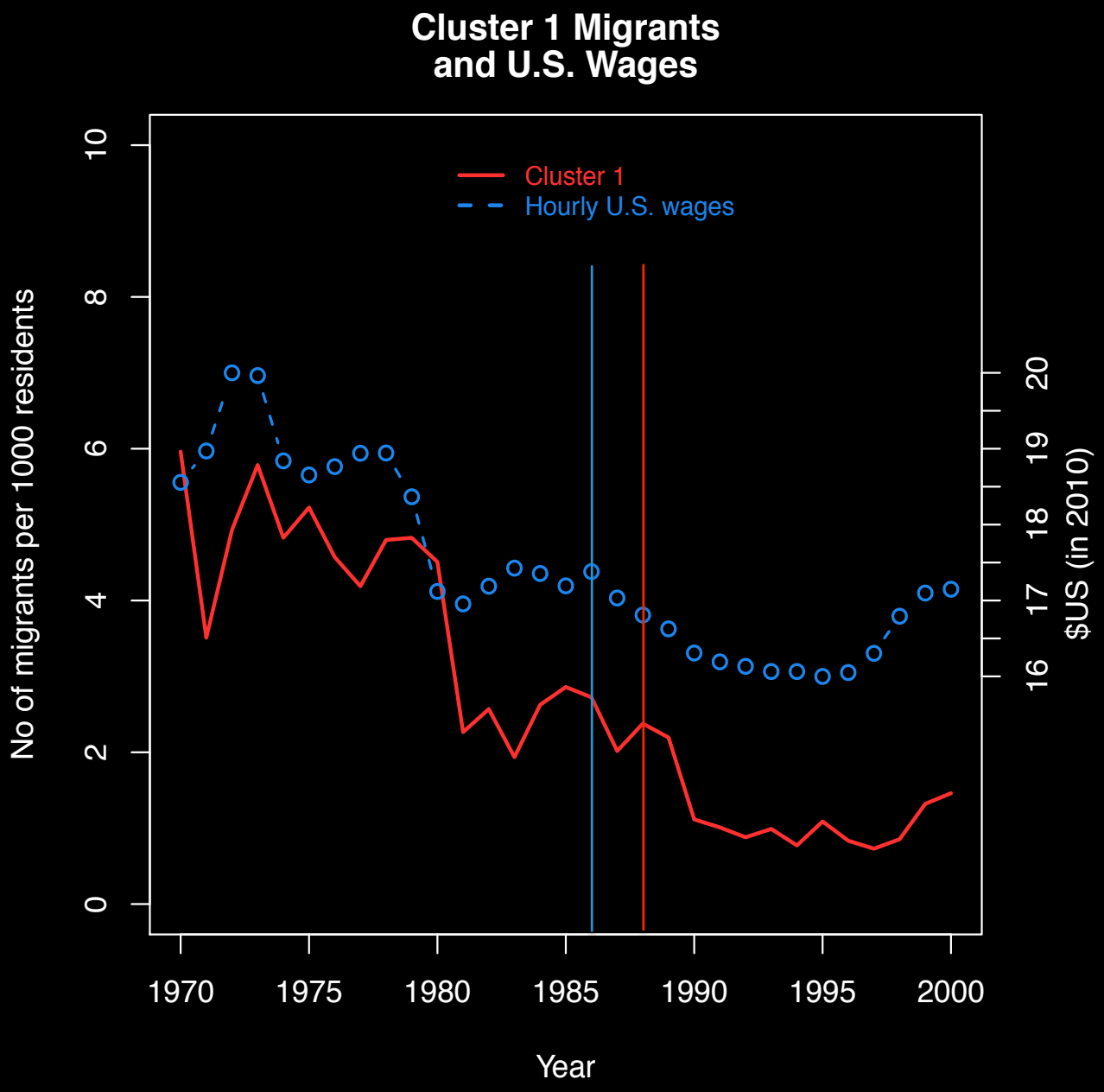
$R^2 = 0.9$


Cluster 1 Migrants
and U.S. Wages
Cluster 1
Hourly U.S. wages
No of migrants per 1000 residents
0
2
4
6
8
10
16
17
18
19
20
$US (in 2010)
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 1 Migrants
and U.S. Wages
Cluster 1
Hourly U.S. wages
No of migrants per 1000 residents
0
2
4
6
8
10
16
17
18
19
20
$US (in 2010)
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 1 Migrants
and U.S. Wages
Cluster 1
Hourly U.S. wages
No of migrants per 1000 residents
0
2
4
6
8
10
16
17
18
19
20
$US (in 2010)
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 1 Migrants
and U.S. Wages
Cluster 1
Hourly U.S. wages
No of migrants per 1000 residents
0
2
4
6
8
10
16
17
18
19
20
$US (in 2010)
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 1 Migrants
and U.S. Wages
Cluster 1
Hourly U.S. wages
No of migrants per 1000 residents
0
2
4
6
8
10
16
17
18
19
20
$US (in 2010)
1970
1975
1980
1985
1990
1995
2000
Year

Are different clusters responding to different macro-level conditions?

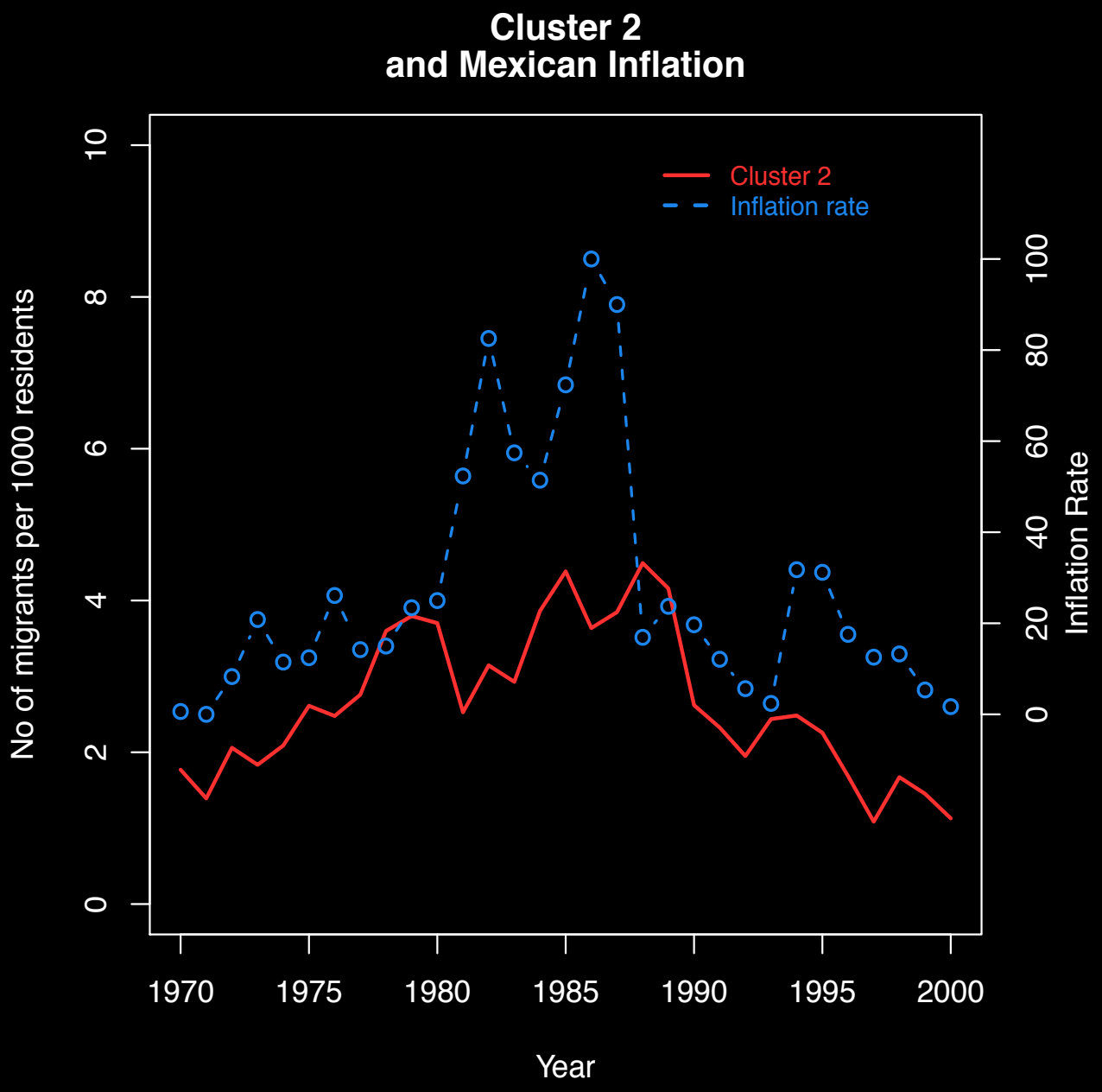
Cluster 1 ~ Neoclassical model

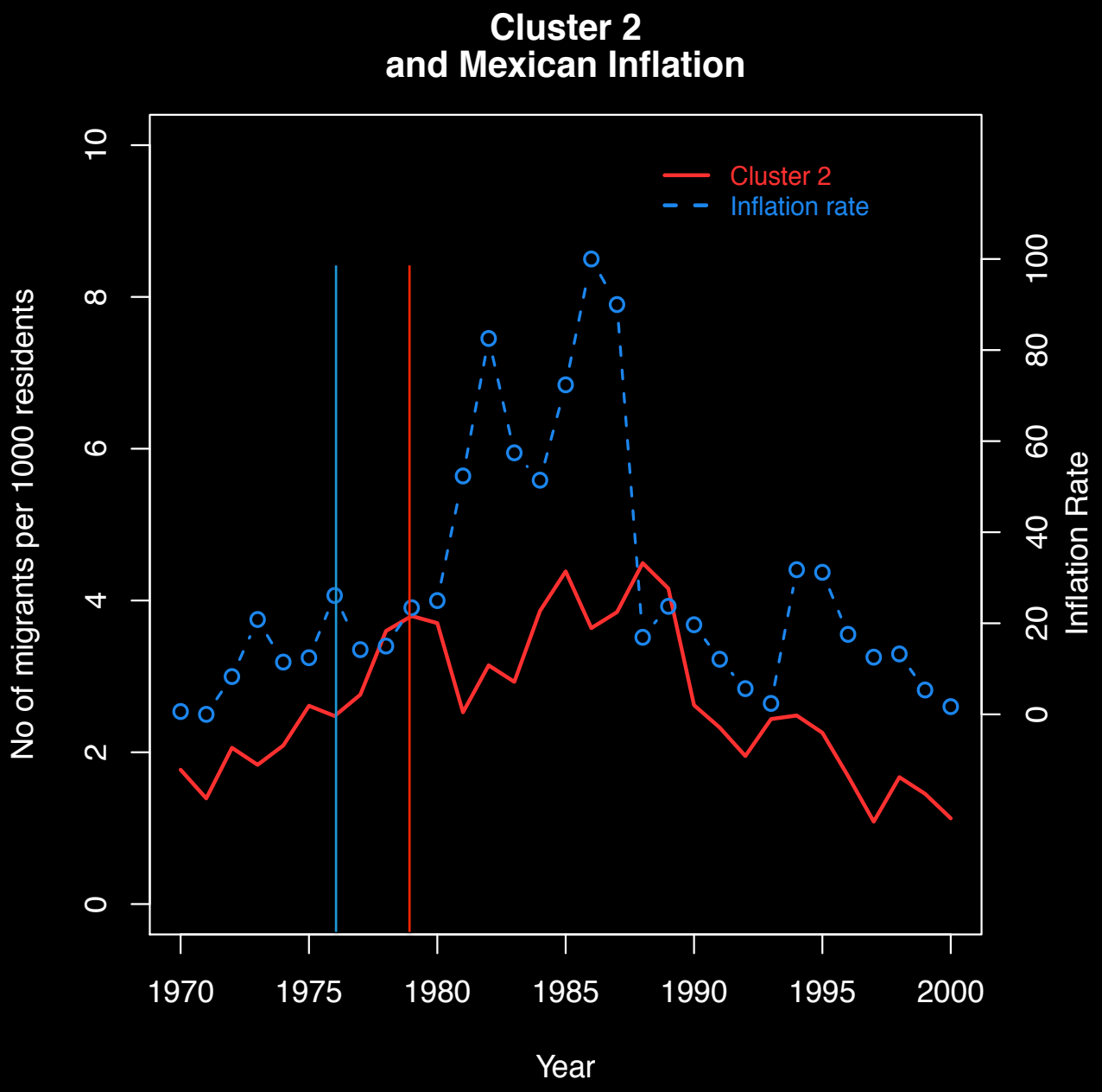
Are different clusters responding to different macro-level conditions?

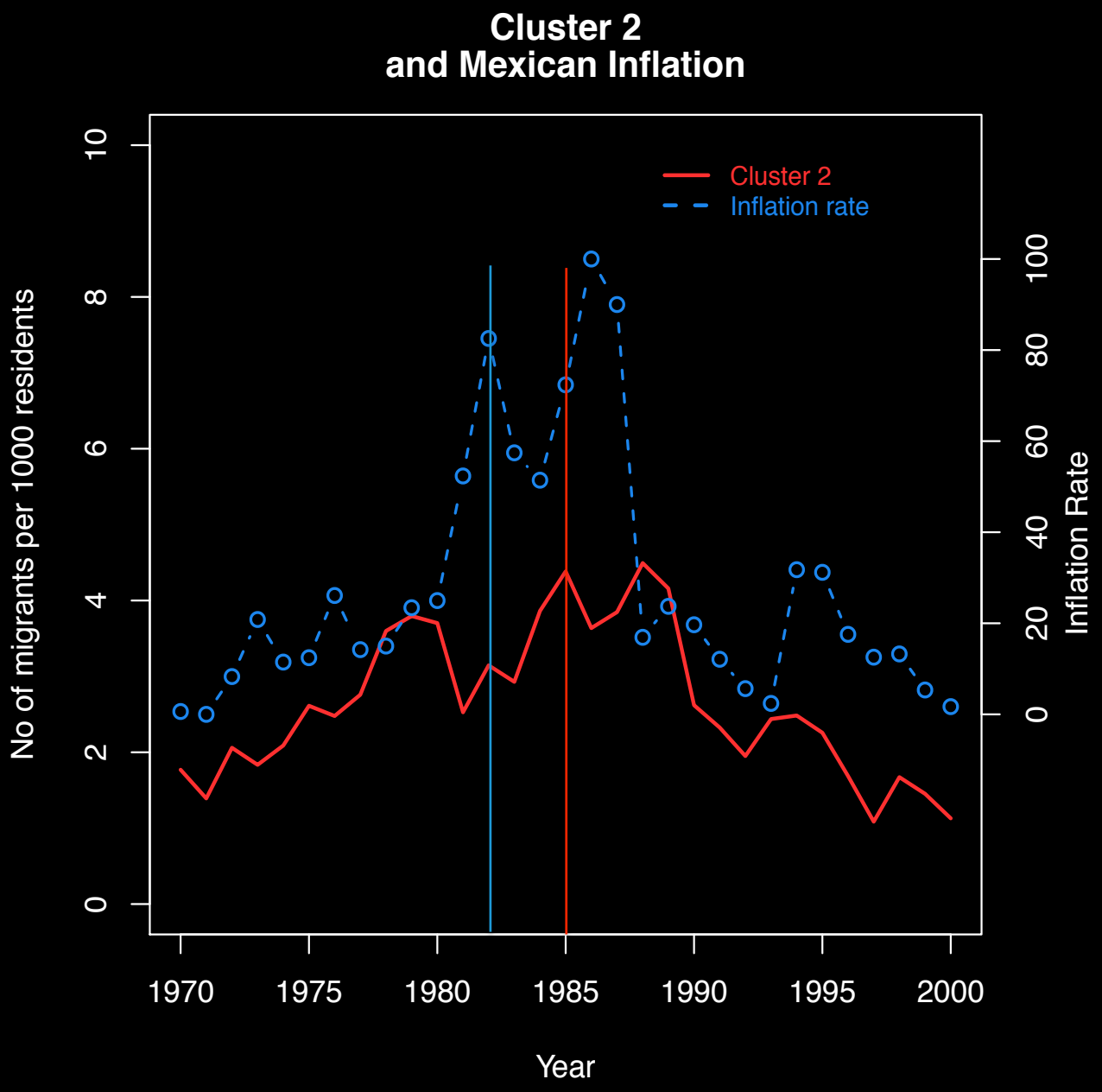
Cluster 2

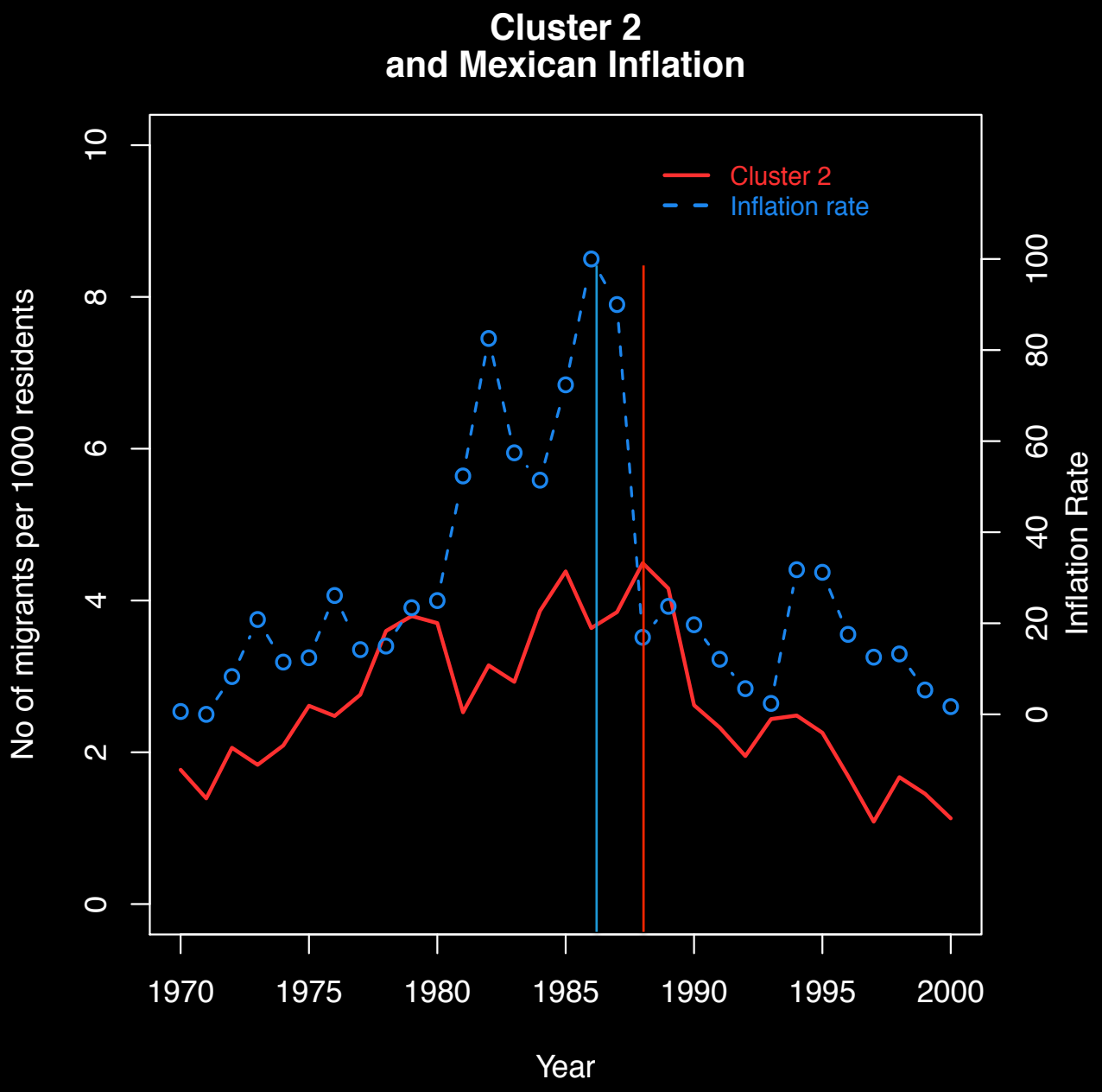
MX inflation	+
BPE budget	-

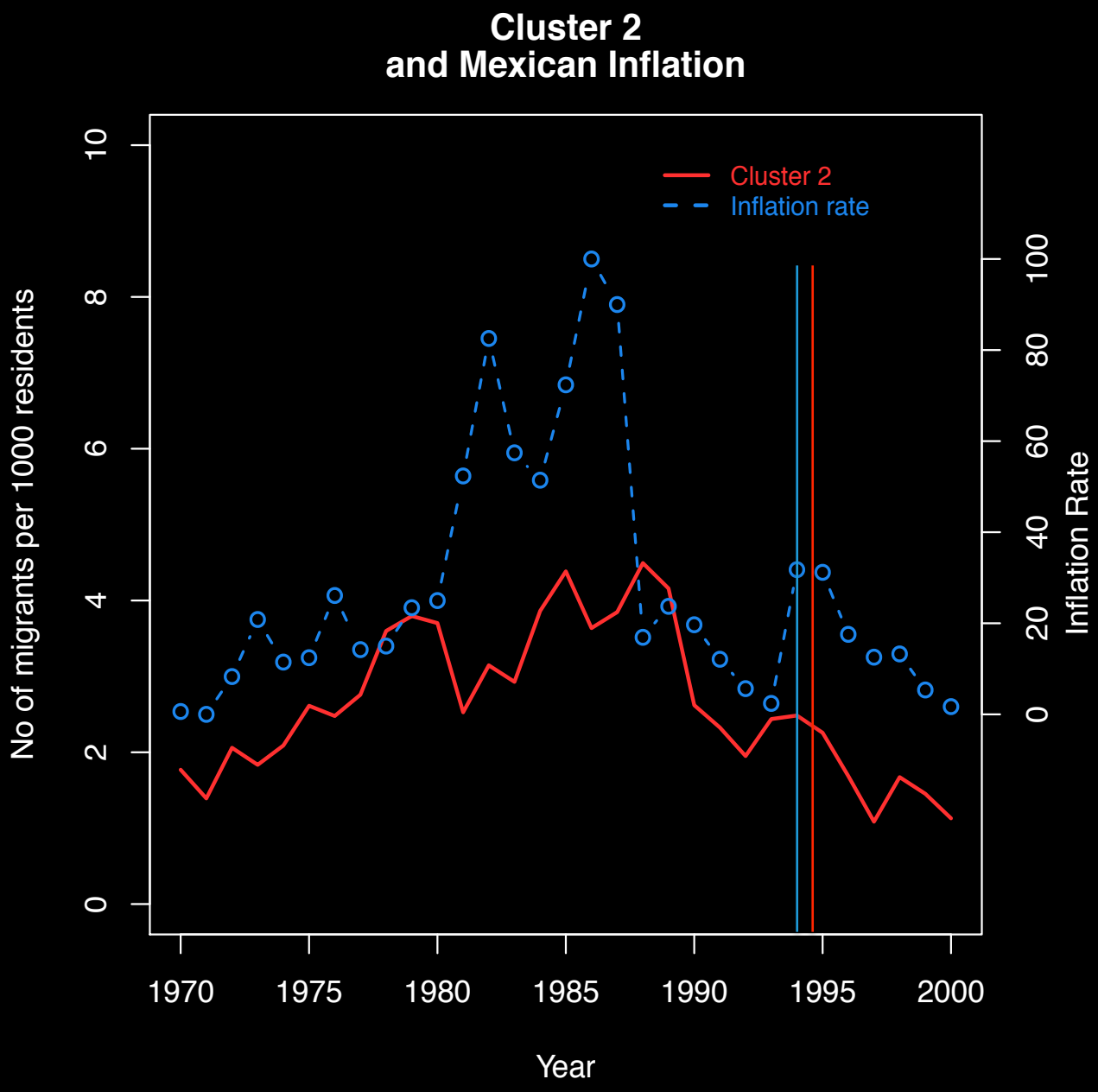
$R^2 = 0.5$


Cluster 2
and Mexican Inflation
Cluster 2
Inflation rate
No of migrants per 1000 residents
Inflation Rate
Year
0
2
4
6
8
10
0
20
40
60
80
100
1970
1975
1980
1985
1990
1995
2000


Cluster 2
and Mexican Inflation
Cluster 2
Inflation rate
No of migrants per 1000 residents
Inflation Rate
Year
0
2
4
6
8
10
0
20
40
60
80
100
1970
1975
1980
1985
1990
1995
2000


Cluster 2
and Mexican Inflation
Cluster 2
Inflation rate
No of migrants per 1000 residents
Inflation Rate
Year
0
2
4
6
8
10
0
20
40
60
80
100
1970
1975
1980
1985
1990
1995
2000


Cluster 2
and Mexican Inflation
Cluster 2
Inflation rate
No of migrants per 1000 residents
Inflation Rate
Year
0
2
4
6
8
10
0
20
40
60
80
100
1970
1975
1980
1985
1990
1995
2000


Cluster 2
and Mexican Inflation
Cluster 2
Inflation rate
No of migrants per 1000 residents
Inflation Rate
Year
0
2
4
6
8
10
0
20
40
60
80
100
1970
1975
1980
1985
1990
1995
2000

Are different clusters responding to different macro-level conditions?

Cluster 1 ~ Neoclassical model

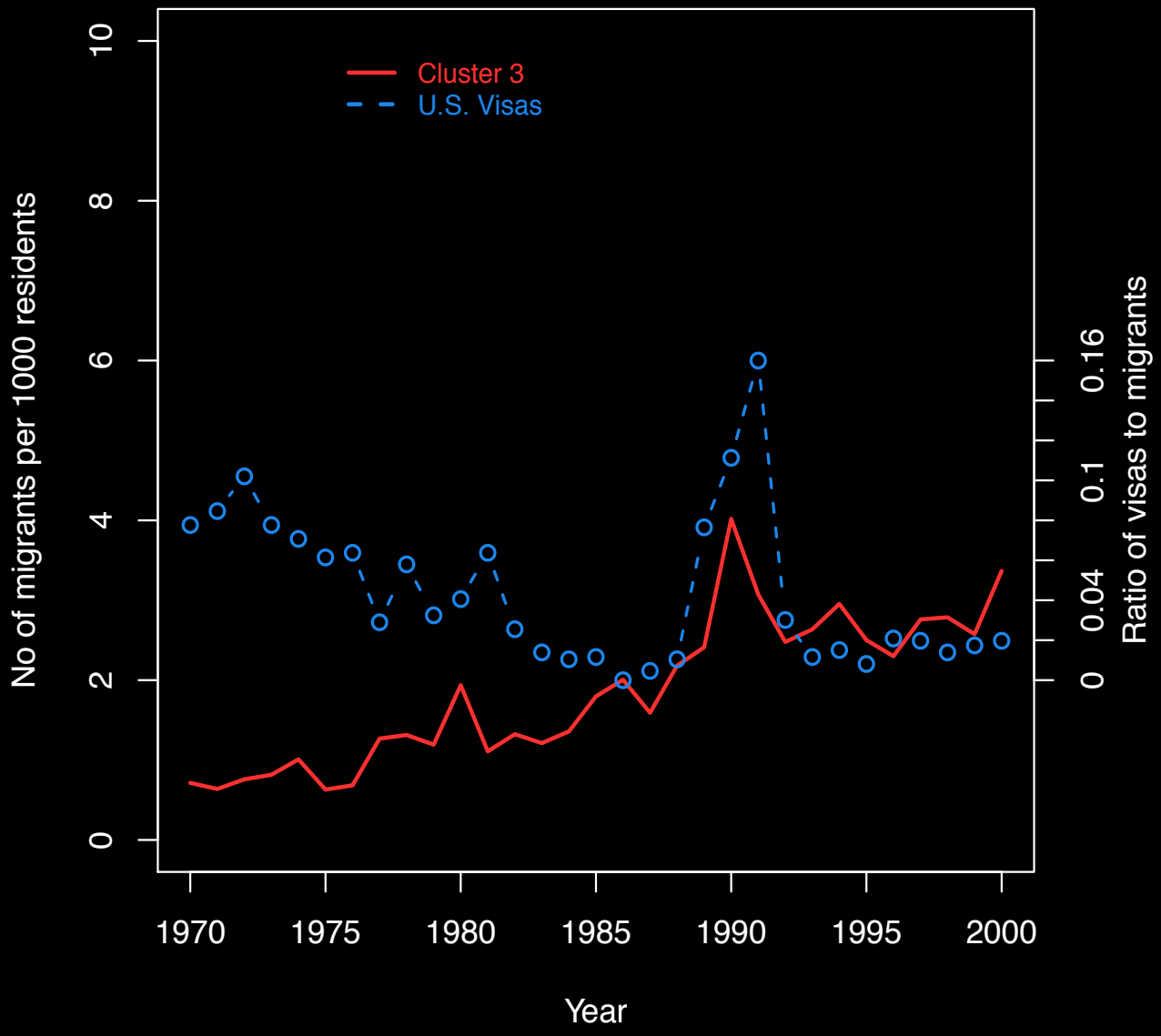
Cluster 2 ~ New economics model

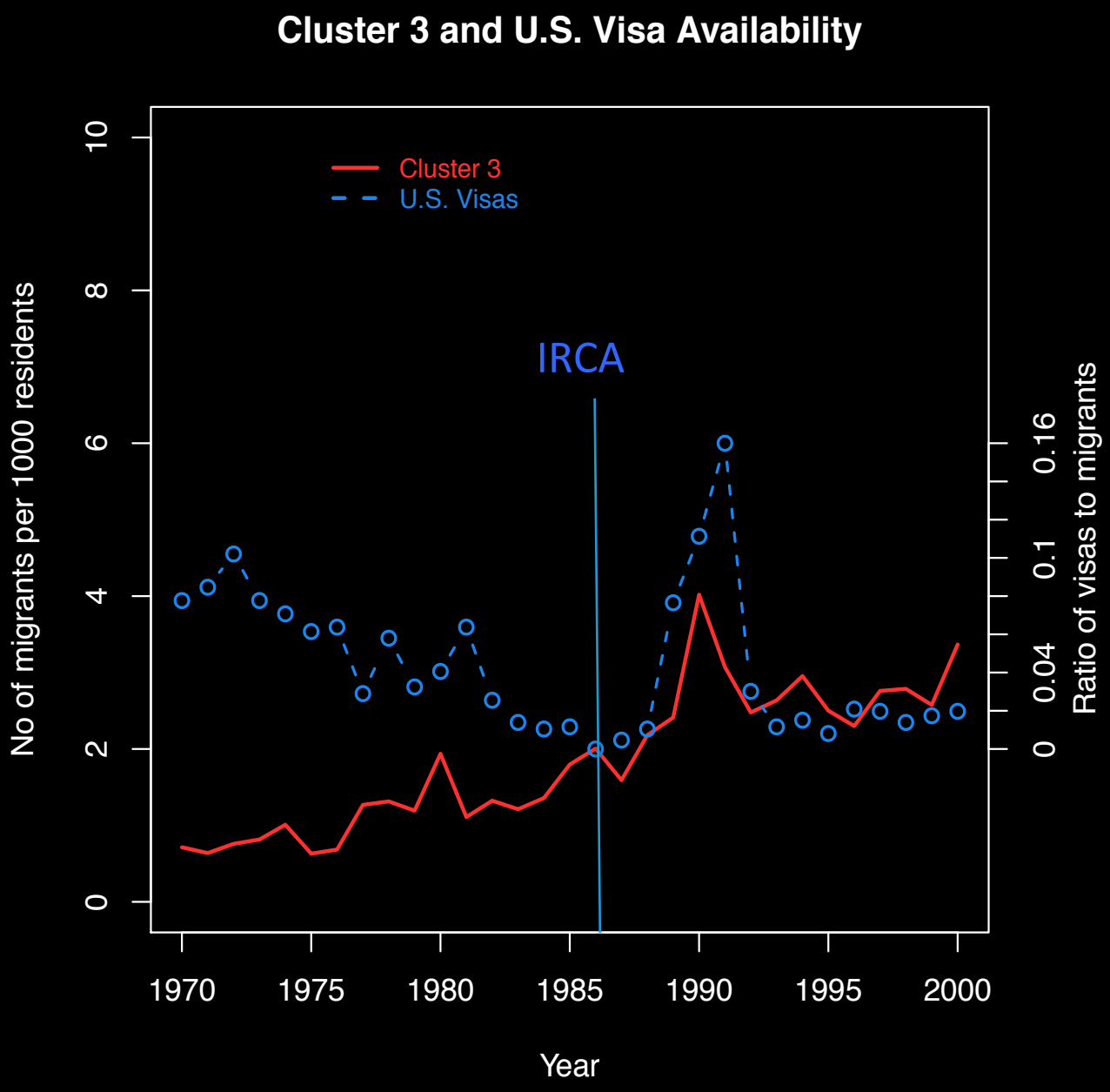
Are different clusters responding to different macro-level conditions?

Cluster 3

US visas for MX	+
MX-US trade	+

$R^2 = 0.8$


Cluster 3 and U.S. Visa Availability
Cluster 3
U.S. Visas
No of migrants per 1000 residents
Ratio of visas to migrants
0
2
4
6
8
10
0
0.04
0.1
0.16
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 3 and U.S. Visa Availability
Cluster 3
U.S. Visas
IRCA
No of migrants per 1000 residents
Ratio of visas to migrants
0
2
4
6
8
10
0
0.04
0.1
0.16
1970
1975
1980
1985
1990
1995
2000
Year

Are different clusters responding to different macro-level conditions?

Cluster 1 ~ Neoclassical model

Cluster 2 ~ New economics model

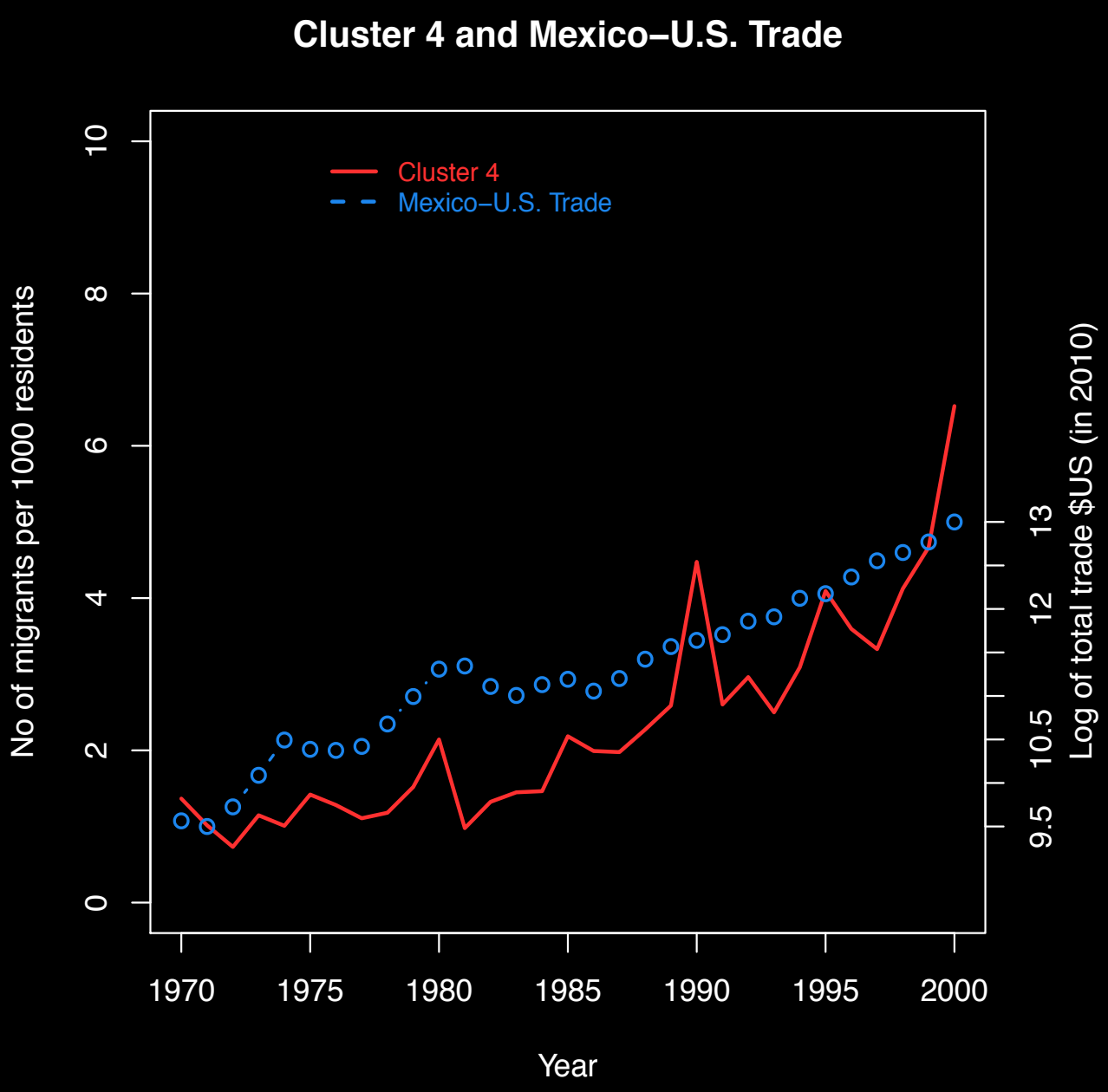
Cluster 3 ~ Cumulative causation model

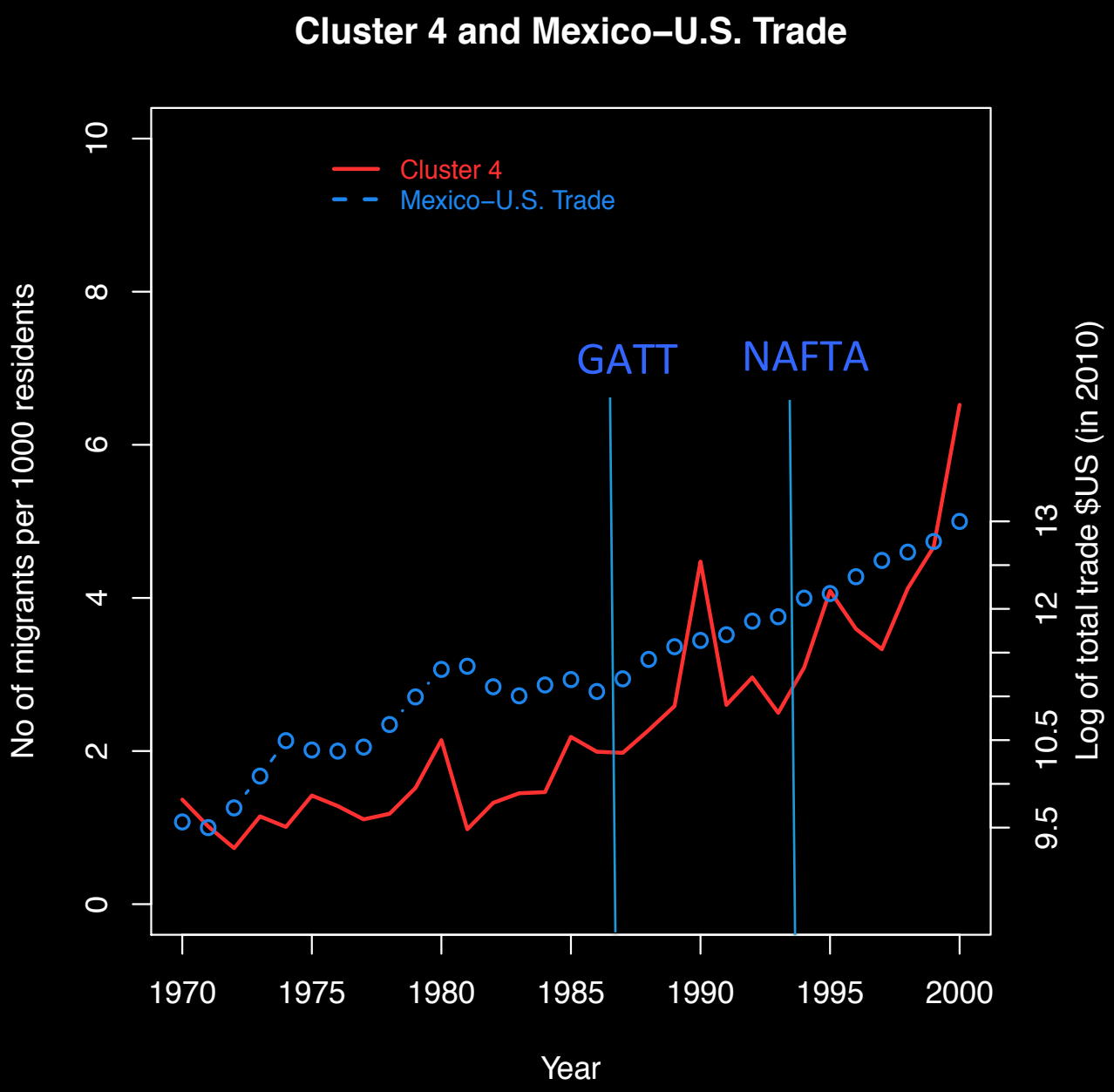
Are different clusters responding to different macro-level conditions?

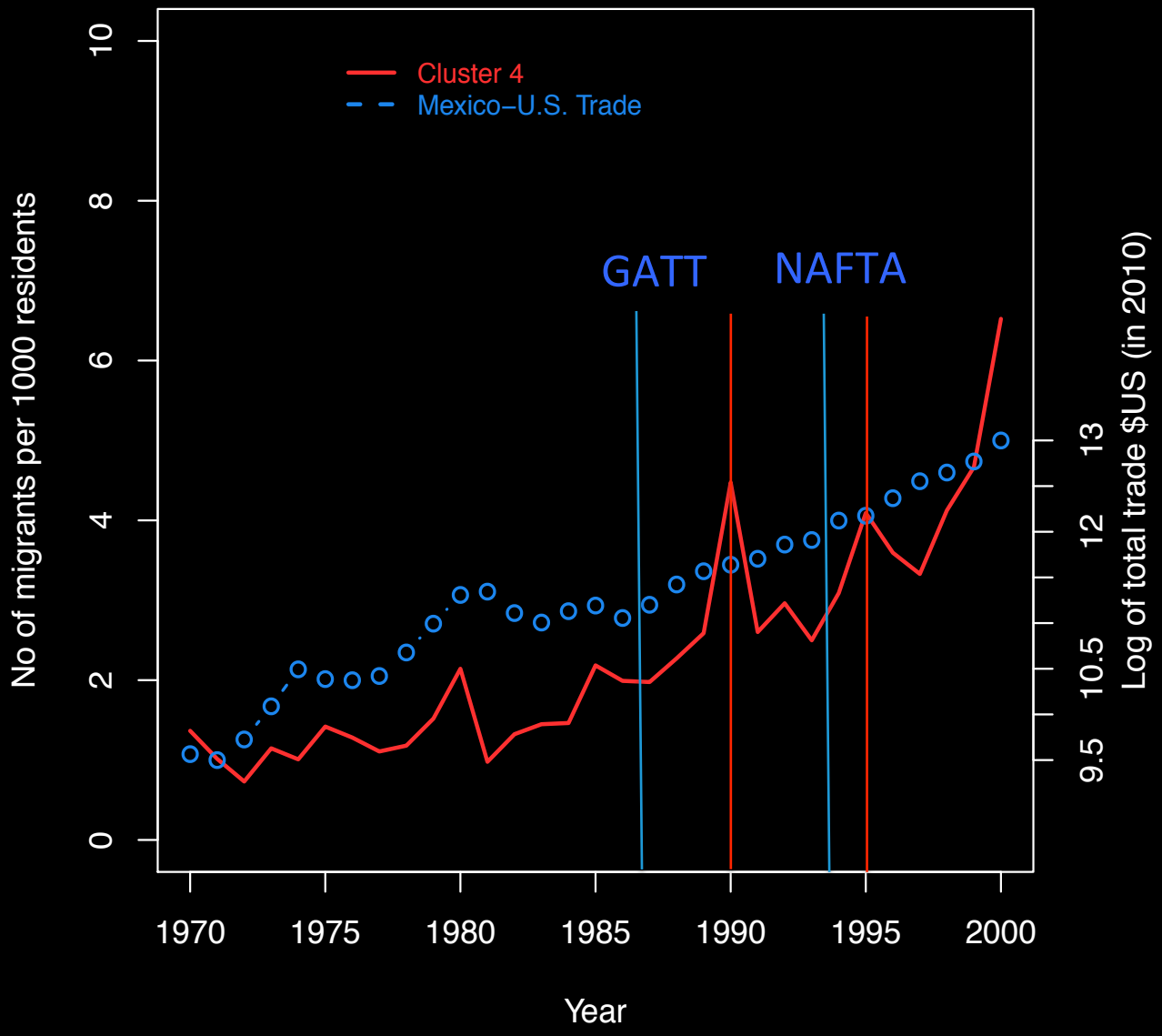
Cluster 4

MX-US trade	+
BPE budget	+

$R^2 = 0.8$


Cluster 4 and Mexico–U.S. Trade
Cluster 4
Mexico–U.S. Trade
No of migrants per 1000 residents
0
2
4
6
8
10
9.5
10.5
12
13
Log of total trade $US (in 2010)
1970
1975
1980
1985
1990
1995
2000
Year


Cluster 4 and Mexico–U.S. Trade
Cluster 4
Mexico–U.S. Trade
GATT
NAFTA
No of migrants per 1000 residents
Log of total trade $US (in 2010)
Year
0
2
4
6
8
10
9.5
10.5
12
13
1970
1975
1980
1985
1990
1995
2000


Cluster 4 and Mexico–U.S. Trade
Cluster 4
Mexico–U.S. Trade
GATT
NAFTA
No of migrants per 1000 residents
0
2
4
6
8
10
Log of total trade $US (in 2010)
9.5
10.5
12
13
1970
1975
1980
1985
1990
1995
2000
Year

Are different clusters responding to different macro-level conditions?

Cluster 1 ~ Neoclassical model

Cluster 2 ~ New economics model

Cluster 3 ~ Cumulative causation model

Cluster 4 ~ World systems model (?)

Where do we go from here?

"...one of two or three cardinal problems that social science has not yet come to grips with is precisely this issue of heterogeneity… The ubiquity of heterogeneity means that for the most part we substitute actuarial probabilities for the true individual probabilities, and therefore we generate mainly descriptively accurate but theoretically empty and prognostically useless statistics."(Letter from Otis Dudley Duncan to Yu Xie, 30 July 1996)

"The most important discovery [in microeconomic investigations] was the evidence on the pervasiveness of heterogeneity and diversity in economic life. When a full analysis of heterogeneity in responses was made, a variety of candidate averages emerged to describe the "average" person, and the long-standing edifice of the representative consumer was shown to lack empirical support." (James Heckman, Nobel Memorial Lecture in Economic Sciences, 8 December 2000)