Identifying Neighborhood Effects on Unemployment in the French Case

Florence Goffette-Nagot
GATE - CNRS - Université de Lyon
Claire Dujardin
IWEPS

Workshop "Addressing Methodological Challenges in the Neighbourhood Effects Research"

Nuffield Foundation, London, 22nd March 2019

Motivation

- **Neighborhood composition** (education, unemployment, ...) may impact **individuals' labour market outcomes** through lack of role models, role of social networks in the job search process, ...

$\Rightarrow$ Does living in a deprived neighborhood impact individual's unemployment risk?

- Important in order to better design **public policies** aimed at fighting unemployment in deprived neighbourhoods
  - If there are NE, helping unemployed individuals move to other neighbourhoods could be efficient
  - If not, bringing new jobs into these neighborhoods might be better
  - Are urban renewal policies efficient tools to fight unemployment in these areas-?

Existing literature on neighbourhood effects

- **Mechanisms** behind neighborhood effects:
  see e.g. survey by Gobillon et al., Urban Studies, 2007
- **Existing evaluations**:
  no consistent evidence; clearly depends on the identification strategy and context

Our contribution

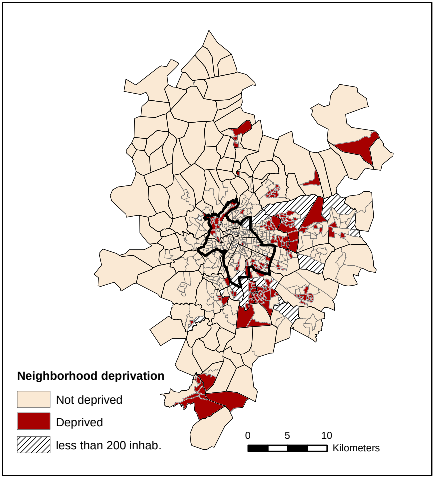
- Two papers:
  - Dujardin C., Goffette-Nagot F., 2010, Neighborhood effects on unemployment? A test *à la* Altonji, *Regional Science and Urban Economics*.
  - Dujardin C., Goffette-Nagot F., 2009, Does public housing occupancy increase unemployment?, *Journal of Economic Geography*.
- Method:
  - Estimate the effect of "**living in deprived neighb.**" or "in public housing" (0/1) in a probit **equation of unemployment** (0/1).
    $\Rightarrow$ probit estimates are biased.
  - Use **two different identification methods** to obtain the **causal impact**
- Results:
  - **No detrimental impact** of public housing and living in a deprived neighbourhood on unemployment

Definition of deprived neighbourhoods in Lyon (1/2)

- Data and sample
  - Urban agglomeration of Lyon
  - French population census, 1999 (1:20th sample, detailed personal and household characteristics)
  - Sample: male heads of couple households aged 19-64 (10,473 indiv.)
- Definition of deprived neighborhoods
  - Basic spatial unit (neighbourhood): about 2000 inhabitants
  - Deprivation continuous index: composite indicator based on a set of **socioeconomic indicators of population** (education, profession, unemployment, foreign nationality, single-mothers)

  ⇒ **Deprivation binary variable**: cut-off value to define the 25% most deprived neighbourhoods

Definition of deprived neighbourhoods in Lyon (2/2)



	Deprived	Others	Total
% Pub. housing	51.9	11.1	21.3
% Unemployed	20.3	9.6	12.3
% Foreigners	23.1	6.9	10.9
% University dip.	10.6	29.3	24.6
% Blue-collars	38.0	16.2	21.7

Does living in one of these neighbourhoods increase per se unemployment probability?

Endogeneity issue: causes

- In the data, we observe a correlation between individual's unemployment probability and neighbourhood composition; three possible reasons:
  - **Reverse causality**: unemployed individuals are more likely to live in DN because of lower housing prices
  - **Sorting**: households' preferences for location may be correlated to preferences affecting outcomes on the labour market
  - **Causal impact of neighbourhood quality** on individual outcomes on the labour market
- Crucial to find a way to estimate the **causal impact**, not biased by the other two mechanisms

Why looking for the causal impact?

- Crucial to be able to disentangle
  - Observing the **simultaneous variation of** DN_i **and** U_i with changes in unobserved characteristics (u_i)
  - Identifying the **causal effect** living in a deprived neighbourhood (DN_i) on the likelihood to be unemployed (U_i)
- Only the causal impact can be considered as valid for any individual and could be obtained for a different population
- To obtain the causal impact:
  compare individuals with the same observed and unobserved characteristics (u_i), in deprived/other neighbourhoods

Endogeneity issue: vocabulary

- Correlation between unobservables, or **selection/sorting based on unobservables**: individuals have unobserved traits (i.e. not observed in the data), that make them locate in deprived/undeprived neighb. and affect their outcome on the labour market
- Probit estimation gives a **naive estimate**: the model is estimated ignoring the correlation of unobservables affecting location and unemployment
- **Unbiased estimate**: reflects the **causal impact** of living in public housing on unemployment
- **Identification strategy**: a way to identify the causal impact

Three possible identification strategies

1. **Take public housing tenants?**
   - Motivation: PH tenants do not choose their housing unit, so its exact location is not correlated with their unobserved characteristics
   - But, not all eligible households apply to PH; moreover, when people are offered a dwelling, they can refuse it if they dislike the location
   - In the end, the location of PH tenants is likely to be correlated with their unobservables (preference for the present, ...)

   $\Rightarrow$ Not a valid identification strategy in the French case

Three possible identification strategies

1. **Take public housing tenants?**: no
2. **Use an instrumental variable method**: find something that produces a **shift in location choices**, but without impacting labour market outcomes: individuals at the margin will shift between deprived/other neighborhoods when this variable changes; allows to compare individuals affected and not affected by this shift
3. **Altonji's method**: Make hypotheses on the **strength of sorting** into locations and look at the impacts on the estimated effect

Three possible identification strategies

1. **Take public housing tenants?**: no
2. **Use an instrumental variable method**: find something that produces a **shift in location choices**, but without impacting labour market outcomes: individuals at the margin will shift between deprived/other neighborhoods when this variable changes; allows to compare individuals affected and not affected by this shift
3. **Altonji's method**: Make hypotheses on the **strength of sorting** into locations and look at the impacts on the estimated effect

Identification strategy 1: Instrumental variables

- **Use an instrumental variable method**: find a variable that impacts the likelihood for households to live in PH or DN, without being correlated to unemployment probability
- **Three different instruments**
- On a national sample: the **share of PH** varies widely across cities in France; it impacts significantly the probability to be in PH, and there is no reason for it to be correlated to individual unobservables
- On Lyon sample:
  - Having **2 children of mixed gender** decreases the probability to have a 3rd child, hence the probability to live in PH (priority given to large families) and therefore in DN
  - **Spouse's workplace**: when the spouse works in the eastern part of Lyon (where most of PH is located), the household is more likely to live in PH and DN

Identification strategy 1: Instrumental variables

- **Use an instrumental variable method**: find a variable that impacts the likelihood for households to live in PH or DN, without being correlated to unemployment probability
- **Three different instruments**
- On a national sample: the **share of PH** varies widely across cities in France; it impacts significantly the probability to be in PH, and there is no reason for it to be correlated to individual unobservables
- On Lyon sample:
  - Having **2 children of mixed gender** decreases the probability to have a 3rd child, hence the probability to live in PH (priority given to large families) and therefore in DN
  - **Spouse's workplace**: when the spouse works in the eastern part of Lyon (where most of PH is located), the household is more likely to live in PH and DN

Identification strategy 1: Instrumental variables

- **Use an instrumental variable method**: find a variable that impacts the likelihood for households to live in PH or DN, without being correlated to unemployment probability
- **Three different instruments**
- On a national sample: the **share of PH** varies widely across cities in France; it impacts significantly the probability to be in PH, and there is no reason for it to be correlated to individual unobservables
- On Lyon sample:
  - Having **2 children of mixed gender** decreases the probability to have a 3rd child, hence the probability to live in PH (priority given to large families) and therefore in DN
  - **Spouse's workplace**: when the spouse works in the eastern part of Lyon (where most of PH is located), the household is more likely to live in PH and DN

Identification strategy 2: Assumptions on selection into locations ("*à la*" Altonji)

Altonji et al.'s method

- The bias is due to correlation between unobservables that influence residential location and unemployment
- We don't know the actual **strength of this correlation** (i.e. sorting into locations) as it is by definition unobserved, but we can **make hypotheses**
- Assuming different correlation levels give different values for the impact of DN on U
    - Assume a zero correlation (probit estimate): higher bound of the estimated NE (i.e. there is no sorting, what we observe is indeed the results of NE)
    - Assume a large correlation (i.e. individuals in DN differ a lot in terms of unobservables impacting labour market outcomes): lower bound of the estimated NE

Identification with instruments - PH impact - First stage

Table 3. Coefficients estimated from probit models of public housing accommodation

	Lyon sample Population Census		
	(1)	(2)	(3)
Probits models of public housing			
Girl+boy	−0.1283 (0.0542)**	−0.1217 (0.0543)**	−0.1289 (0.0544)**
Four children or more		0.3438 (0.0737)***	
Moved within municipality			0.2557 (0.0528)***
Percentage of public housing in the urban unit			
Log likelihood	−1944	−1934	−1932
Pseudo R^2	0.255	0.259	0.259
Number of observations	4849	4849	4849
Test on instruments from GMM estimation of linear probability models			
H0: all instruments zero			
1st stage F-test	5.13	14.41	13.89
[p-value]	(0.024)	(0.000)	(0.000)
H0: instruments orthogonal to error term			
2nd stage overid. test Hansen J	–	0.771	1.466
[p-value]		(0.379)	(0.266)

Identification with instruments - Results for NE

Explained variable: dummy for being unemployed

	All individuals		**>= 2 children**	
	Probit	Bivariate probit	Probit	Bivariate probit
Deprived neighborhood	0.0213***	-0.0329	0.0265***	0.0026
	(0.0063)	(0.0304)	(0.0082)	(0.0306)
Instruments				
Spouse workplace		0.1123***		0.1335***
		(0.0.212)		(0.0281)
Girl-Boy				-0.0231*
				(0.0136)
Tests on instruments				
1st stage F [p-value]		28.57 [0.000]		12.25 [0.000]
Overid. Hansen J [p-value]				0.530 [0.466]
Correlation of residuals		0.301		0.133
		(0.1932)		(0.1802)

Identification with instruments - Results for NE

Explained variable: dummy for being unemployed

	All individuals		**>= 2 children**	
	Probit	Bivariate probit	Probit	Bivariate probit
Deprived neighborhood	0.0213***	-0.0329	0.0265***	0.0026
	(0.0063)	(0.0304)	(0.0082)	(0.0306)
Instruments				
Spouse workplace		0.1123***		0.1335***
		(0.0.212)		(0.0281)
Girl-Boy				-0.0231*
				(0.0136)
Tests on instruments				
1st stage F [p-value]		28.57 [0.000]		12.25 [0.000]
Overid. Hansen J [p-value]				0.530 [0.466]
Correlation of residuals		0.301		0.133
		(0.1932)		(0.1802)

Identification with instruments - Results for NE

Explained variable: dummy for being unemployed				
	All individuals		**>= 2 children**	
	Probit	Bivariate probit	Probit	Bivariate probit
Deprived neighborhood	0.0213***	-0.0329	0.0265***	0.0026
	(0.0063)	(0.0304)	(0.0082)	(0.0306)
Instruments				
Spouse workplace		0.1123***		0.1335***
		(0.0.212)		(0.0281)
Girl-Boy				-0.0231*
				(0.0136)
Tests on instruments				
1st stage F [p-value]		28.57 [0.000]		12.25 [0.000]
Overid. Hansen J [p-value]				0.530 [0.466]
Correlation of residuals		0.301		0.133
		(0.1932)		(0.1802)

Identification *à la* Altonji - Results

1. Sensitivity analysis: how does the estimated effect react to different assumptions on ρ (i.e. intensity of sorting, or correlation of the unobserved parts of the two outcomes)?

$$\begin{cases} U & = & 1(\alpha DN + X'\beta_1 + u_1 > 0) \\ DN & = & 1(X'\beta_2 + u_2 > 0) \\ \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} & \sim & N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & \rho \\ \rho & 0 \end{bmatrix} \right) \end{cases}$$

Hyp. on ρ	0.00	0.05	0.10	0.15	0.20
Result for $\hat{\alpha}$	0.0213***	0.0106*	0.0007	-0.0085*	-0.0172***
	(0.0063)	(0.0058)	(0.0054)	(0.0050)	(0.0045)

Identification *à la* Altonji - Results for NE

2. Selection on observables = selection on unobservables

$$\begin{cases} U &= 1(\alpha DN + X'\beta_1 + u_1 > 0) \\ DN &= 1(X'\beta_2 + u_2 > 0) \\ \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & \rho \\ \rho & 0 \end{bmatrix}\right) \quad \text{with } \rho = \frac{Cov(X'\beta_2, X'\beta_1)}{Var(X'\beta_1)} \end{cases}$$

	Hyp. of equal selection
$\hat{\alpha}$	-0.1321***
$\hat{\rho}$	0.8137

- Observables (education, age, nationality) are important determinants of unemployment $\rightarrow$ extreme hypothesis on the value of ρ

$\Rightarrow$ Negative and unrealistic impact of DN

- Bounds for the causal impact: [-0.1321 (equal sorting) ; 0.0213 (no sorting)]

Identification *à la* Altonji - Results for NE

3. Amount of selection required to entirely explain the probit estimate (in terms of sorting due to unobservables relative to sorting on observables)

$$\begin{cases} y_1 & = & 1(\alpha y_2 + X'\beta_1 + u_1 > 0) \\ y_2 & = & 1(X'\beta_2 + u_2 > 0) \\ \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} & \sim & N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & \rho \\ \rho & 0 \end{bmatrix}\right) \end{cases}$$

$$\frac{E(u_1|y_2{=}1) - E(u_1|y_2{=}0)}{var(u_1)} = \lambda \frac{E(X'\beta_1|y_2{=}1) - E(X'\beta_1|y_2{=}0)}{var(X'\beta_1)}$$

Compute λ so that $\hat{\alpha} = 0$

Result $\lambda = 5.5\%$ is enough to explain the naive probit effect

i.e. a low level of sorting is enough to produce the estimated effect

Conclusion

- Different identification methods. A choice to be made **depending on the context**.
- **Altonji's method**: no single impact, but gives bounds for the causal estimate; useful when no other identification method available, or to confirm the result of another method
- **Heterogenous treatment effects**: NE are likely to be more or less strong depending on individuals' characteristics. E.g. the impact might depend on individual's educational level.
- Instrumental variable estimates give a value for the **Local Average Treatment Effect (LATE)**: the impact for individuals who are indeed likely to be affected by the instrument.
  - In our case: families with children, and not singles or elderly.
- Our results tend to show that in France, residential situation has **no impact on unemployment probability**. This does not rule out any other impact (crime, health, ...).