# 2009/56

■

# Neighbourhood effects and endogeneity issues

Claire Dujardin, Dominique Peeters
and Isabelle Thomas

## CORE

# DISCUSSION PAPER

# Neighbourhood effects and endogeneity issues

Claire DUJARDIN[1], Dominique PEETERS[2]
and Isabelle THOMAS [3]

September 2009

## Abstract

A recent body of research suggests that the spatial structure of cities might influence the socioeconomic characteristics and outcomes of their residents. In particular, the literature on neighbourhood effects emphasizes the potential influence of the socioeconomic composition of neighbourhoods in shaping individual's behaviours and outcomes, through social networks, peer influences or socialization effects. However, empirical work still has not reached a consensus regarding the existence and magnitude of such effects. This is mainly because the study of neighbourhood effects raises important methodological concerns that have not often been taken into account. Notably, as individuals with similar socio-economic characteristics tend to sort themselves into certain parts of the city, the estimation of neighbourhood effects raises the issue of location choice endogeneity. Indeed, it is difficult to distinguish between neighbourhood effects and correlated effects, i.e. similarities in behaviours and outcomes arising from individuals having similar characteristics. This problem, if not adequately corrected for, may yield biased results.

In the first part of this paper, neighbourhood effects are defined and some methodological problems involved in measuring such effects are identified. Particular attention is paid to the endogeneity issue, giving a formal definition of the problem and reviewing the main methods that have been used in the literature to try to solve it. The second part is devoted to an empirical illustration of the study of neighbourhood effects, in the case of labour-market outcomes of young adults in Brussels. The effect of living in a deprived neighbourhood on the unemployment probability of young adults residing in Brussels is estimated using logistic regressions. The endogeneity of neighbourhood is addressed by restricting the sample to young adults residing with their parents. Then, a sensitivity analysis is used to assess the robustness of the results to the presence of both observed and unobserved parental covariates.

---

[1] F.R.S.-FNRS and Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: Claire.dujardin@uclouvain.be
[2] Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: Dominique.peeters@uclouvain.be. This author is also member of ECORE, the association between CORE and ECARES.
[3] F.R.S.-FNRS and Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium. E-mail: isabelle.thomas@uclouvain.be. This author is also member of ECORE, the association between CORE and ECARES.

## 1. Introduction

The last twenty-five years have seen a rising interest among economists, social scientists and geographers in the study of the way in which neighbourhood context may affect individual behaviours and outcomes. The work of sociologist W.J. Wilson has been particularly influential in suggesting that social influences in the neighbourhood, i.e. neighbourhood effects, could be part of the explanation for the social problems experienced by poor inner-city residents in American metropolitan areas.

> *...in a neighbourhood with a paucity of regularly employed families and with the overwhelming majority of families having spells of long term joblessness, people experience a social isolation that excludes them from the job network system that permeates other neighbourhoods and that is so important in learning about or being recommended for jobs... In such neighbourhoods, the chances are overwhelming that children will seldom interact on a sustained basis with people who are employed or with families that have a steady breadwinner. The net effect is that joblessness, as a way of live, takes on a different meaning...*

> Wilson, *The Truly Disadvantaged* (1987, p. 57)

Studies aiming at estimating neighbourhood effects are numerous and have been the focus of several extensive surveys (see, among others, Jencks and Mayer, 1990; Ellen and Turner, 1997; Sampson et al, 2002; Dietz, 2002; Durlauf, 2004). These studies focus on a wide variety of outcomes such as teenagers' educational attainment and school attendance, delinquency, drug consumption, out-of-wedlock pregnancy, as well as labour-market outcomes (unemployment and welfare participation) and health status (for example, well-being and mental health).

Despite the bulk of empirical studies, there is still considerable debate over the existence and magnitude of neighbourhood effects. Some authors consider that neighbourhoods play a significant role in explaining individual outcomes while others are convinced that their role –if any– are of marginal importance compared to that of personal characteristics and backgrounds. For example, Jencks and Mayer (1990) conclude their survey on the role of neighbourhood effects in shaping children and adolescents' behaviours by saying that "the literature we reviewed does not […] warrant any strong generalizations about neighbourhood effects" (p. 176).

According to several authors (Ginther et al, 2000; Dietz, 2002), the primary reason for this lack of consensus is the great diversity regarding the methods, data and variables used to test for the existence of neighbourhood effects. Moreover, the estimation of neighbourhood effects is a difficult task. Recent studies have highlighted the existence of important methodological problems inherent to the estimation of such effects, which have rarely been accounted for in most earlier empirical works (Manski, 1993; Durlauf, 2004; Blume and Durlauf, 2006). Of paramount importance is the endogeneity bias that results from the self-selection of individuals into neighbourhoods (Blume and Durlauf, 2006). Indeed, trying to explain individual outcomes by neighbourhood characteristics in a simple regression analysis does not lead to concluding results as residential locations are not exogenously determined. On the contrary, individuals having similar characteristics tend to sort themselves in some parts of the urban space. There is thus a two-way causality: on the one hand, residential location influences individual socioeconomic outcomes, and on the other hand, individual outcomes influence the choice of a residential location.

Standard econometric methods are unable to distinguish between two-way causality and may consequently yield biased results.

This paper seeks to contribute to the field by defining neighbourhood effects and giving an overview of existing solutions to overcome endogeneity problems. The first section gives a conceptual definition for what lies behind the generic term of "neighbourhood effects" in order to precisely define what this paper is concerned with. The second section briefly lists some of the more recurrent methodological problems that have been encountered by most empirical work on neighbourhood effects. The third section then focuses on the most prominent problem in this field: the endogeneity associated with residential location. It provides a precise formulation of the endogeneity issue and reviews critically the various methods that have been proposed to solve it. The rest of the paper is devoted to an empirical application of the study of neighbourhood effects. Instead of trying to solve the endogeneity issue, the importance of endogeneity biases is assessed through a sensitivity analysis, in order to evaluate the robustness of the results. This method is illustrated with estimates of neighbourhood effects on labour-market outcomes of young adults residing in Brussels.

## 2. Neighbourhood effects: definition and identification issues

There are numerous reasons why residential location should matter in explaining individual behaviours and outcomes (see Ellen and Turner, 1997, Jencks and Mayer, 1990 for extensive surveys). For example, *peer influences* refer to a contagion effect in which the propensity to adopt a socially-deviant behaviour (like dropping out of school, consuming drugs or being unemployed) depends critically on the proportion of peers exhibiting the same behaviour in a community (this is known as the "epidemic theory of ghettos" developed by Crane, 1991). *Socialization* or *role model effects* refer to the influence on the behaviour of teenagers of the socioeconomic success of adults in their neighbourhood of residence, those serving as models to which young people can identify and for what they may aspire to become (Wilson, 1987). The density and composition of *social networks* inside the residential neighbourhood may also influence individual outcomes by conditioning the quantity and quality of information available to individuals regarding access to social services and economic opportunities (Reingold, 1999). Finally, the *physical disconnection* and *isolation* may influence accessibility to services or economic opportunities (see for example the spatial mismatch hypothesis in which distance to job locations explains the high unemployment rates of American inner-city black residents; Ihlanfeldt and Sjoquist, 1998).

These few mechanisms all point to a more general concept, which has generically been labelled "neighbourhood effects" or "contextual effects". However, they cover very different types of residential location influences. It is thus helpful to give some conceptual definitions and to identify what kind of effects will be studied in this paper.

### 2.1. Spatial versus neighbourhood effects

 First, the effects of residential location can be decomposed into pure spatial effects and neighbourhood effects. *Spatial effects* are pure locational effects and refer to the influence on individual outcomes of residing in a particular location in the city, which may give some advantages and disadvantages in terms of accessibility to economic and social opportunities located in the metropolitan area. Clearly, the spatial mismatch hypothesis which postulates that the spatial disconnection between residence and job locations

might negatively affect individual labour-market outcomes is an example of such a spatial effect (Ihlanfeldt and Sjoquist, 1998). On the contrary, *neighbourhood effects* refer to the effects of belonging to a group on the behaviours and socioeconomic outcomes of individuals (Dietz, 2002), independently of the geographical location of the group within the city. Neighbourhood effects refer in general to the effects of belonging to a group defined on a geographical basis. In this case, neighbourhood effects concern influences on an individual's behaviour and outcomes due to the characteristics and behaviours of his/her neighbours.

Although spatial effects and neighbourhood effects are not mutually exclusive, they have often been considered separately. Therefore and for the sake of simplicity, pure spatial effects will be left aside from now and the remainder of this paper will consider only neighbourhood effects. However, most of the methodological problems that will be highlighted in the next sections also apply to pure spatial effects.

## 2.2. Endogenous versus contextual effects

In a widely cited article, Manski (1993) identifies two types of neighbourhood effects (or social interactions, following its own terminology): *endogenous effects*, whereby the propensity of an individual to behave in some way varies with the average behaviour of the group, and *contextual effects* whereby the propensity of an individual to behave in some way varies with the average background characteristics of the group. The distinction between endogenous and contextual effects is subtle as their definitions differ only by one word: a person's behaviour (or choice) is influenced either by the *behaviour* (endogenous effect) or by the *background characteristics* (contextual effect) of those in his/her group. To clarify, consider the example of school achievement. There is an endogenous effect if a teenager's achievement is influenced by the average achievement of his/her peers (i.e. other students in his/her school or neighbourhood). There is a contextual effect if his achievement is influenced by the socioeconomic composition of his neighbourhood, for example by the employment status observed among adults in the neighbourhood, which may provide role models and affect teenagers' aspirations. The relevance of this distinction lies in the fact that endogenous effects generate a social multiplier: acting on an individual's behaviour does not only influence this sole individual, but also exerts an influence on the aggregate behaviour of the group. For example, if a teenager's school achievement increases with the average achievement of the students in the school, then tutoring some students in the school does not only help the tutored students but indirectly helps all students in the school. Contextual effects do not generate such social multipliers (Manski, 1993).

Endogenous and contextual effects may be formalized using the following equation (adapted from Manski, 1993):

$$y_i = \alpha + \beta' X_i + \theta y^e_{n(i)} + \eta' X^e_{n(i)} + \varepsilon_i \qquad [1]$$

where $y_i$ is the outcome of interest of individual $i$ (for example, school achievement), $X_i$ is a vector of individual characteristics (gender, age, family characteristics), $y^e_{n(i)}$ is the mean outcome of those individuals in the neighbourhood $n$ in which individual $i$ resides (mean educational attainment of peers), $X^e_{n(i)}$ is a vector of variables describing neighbourhood composition (including mean values of individual characteristics $X_i$ such as mean parental outcome), and $\varepsilon_i$ is a classical error term. The parameter $\theta$, when significantly different from zero, measures an endogenous effect, while $\eta$ measures a contextual effect.

## 2.3. Neighbourhood versus correlated effects

Besides endogenous and contextual effects (which will be labelled together neighbourhood effects), Manski (1993) identifies a third explanation for the fact that individuals belonging to the same group have similar outcomes: *correlated effects*. In correlated effects, individuals belonging to the same group (in our case to the same neighbourhood) exhibit a similar behaviour simply because they have similar unobserved individual characteristics, and not because of the influence of the group's behaviour and composition. Taking the example of labour-market outcomes, individuals residing in the same neighbourhood may have high unemployment probabilities because living in a neighbourhood with a high proportion of unemployed or low-educated workers generates pervasive effects such as peer effects or poor social networks (a neighbourhood effect). On the contrary, this might simply reflect the fact that individuals living in the same neighbourhood share common unobserved factors that are detrimental in finding a job, such as low ability (a correlated effect).

The following equation amends equation 1 in order to reflect correlated effects:

$$y_i = \alpha + \beta' X_i + \theta y^e_{n(i)} + \eta' X^e_{n(i)} + v_{n(i)} + \varepsilon_i \qquad [2]$$

where the error term is decomposed in two parts, one is the classical error term $\varepsilon_i$ reflecting unobserved characteristics which are peculiar to individual $i$, and the second reflects unobserved characteristics that individual $i$ shares in common with other individuals in his neighbourhood $v_{n(i)}$.

Endogenous and contextual effects both express an influence of the social environment on the behaviours of individuals (i.e. group/neighbourhood matters), whereas correlated effects express a non-social phenomenon (Manski, 1995). Therefore, distinguishing between neighbourhood effects (i.e. endogenous and/or contextual effects) on the one hand and correlated effects on the other hand is important for the design of social policies. Indeed, if neighbourhood effects exist, policies which aim to achieve a more even distribution of individuals across neighbourhoods (for example, by relocating some categories of residents in more socio-economically diverse neighbourhoods) may have an impact on individual outcomes.

## 2.4. Identification issues

Distinguishing between endogenous, contextual and correlated effects using standard observational data is not straightforward and two identification issues have been highlighted in the theoretical literature on social interaction and neighbourhood effects (Blume and Durlauf, 2006).

### 2.4.1. The reflection problem

The first issue pertains to the difficulty in distinguishing between endogenous and contextual effects and is named the reflection problem (Manski, 1993). Indeed, when researchers observe a correlation between an individual outcome $y_i$ and the average outcome in a neighbourhood $y^e_{n(i)}$, they cannot determine whether this correlation is due to the causal influence of aggregate outcomes or to the fact that aggregate outcomes simply reflect the role of average background characteristics in influencing individuals $X^e_{n(i)}$. In very simple terms, the average background characteristics influence average outcomes, which in turn affect individual outcome. In this framework, identification amounts to distinguishing between the direct effect of average background characteristics on individual outcome (the contextual effect) and the indirect effect

of background characteristics, as reflected through the endogenous effect generated by the average outcome of the group (Durlauf, 2004).

Identification of endogenous and contextual effects relies on strong assumptions regarding how the characteristics of the individuals vary with the characteristics of the group, and these assumptions may not be credible according to the nature of the data (see Blume and Durlauf, 2006, for a useful synthesis of formal conditions under which statistical identification is possible in the case of basic linear models as well as binary choice models). Very few studies try to properly disentangle endogenous and contextual effects (Ioannides and Zabel, 2008 is a recent example). Most empirical studies either estimate one of the two effects, positing the assumption that the other is absent, or instead estimate an aggregate of both effects. The same caveat applies to this study, its goal being to prove the existence of neighbourhood effects, whether they arise from endogenous or contextual effects.

### 2.4.2. *The self-selection or endogeneity issue*

The second problem pertains to identifying neighbourhood effects (i.e. contextual and endogenous effects taken together) and disentangling these from correlated effects (i.e. similarities in behaviours and outcomes arising because of unobserved characteristics shared by individuals in the neighbourhood). Correlated effects arise because individuals are not randomly distributed across the urban space. On the contrary, individuals sort themselves into neighbourhoods on the basis of their personal and family background characteristics (for example, income) and some of these characteristics also influence the outcome of interest. These background characteristics are either observed by the researcher, and might be controlled for (i.e. included in $X_i$), or unobserved. Because of these *unobserved* characteristics, distinguishing neighbourhood effect estimates from any correlated effects is difficult. As will become clear later in this paper, if correlated effects are ignored, estimated neighbourhood effects will be biased (Dietz, 2002; Durlauf, 2004). This problem is referred to as the endogeneity issue or the self-selection issue in the literature, as it arises from the fact that individuals choose their neighbourhood of residence (i.e. "self-select" into neighbourhoods). It is important to note that "endogeneity" is used here in its econometric sense, without any relation to Manski's endogenous effect.

The first set of studies on neighbourhood effects completely ignored self-selection issues (see for example Datcher, 1982). However, more recent studies have paid particular attention to developing strategies for coping with this problem and disentangling neighbourhood effects from correlated effects (see for example Dietz, 2002, and Durlauf, 2004, for useful reviews). The purpose of this paper is to review these strategies and to highlight their respective advantages and shortcomings. Doing this, we deliberately ignore questions relating to the distinction of underlying mechanisms through which neighbourhood effects operate (for example the distinction between endogenous and contextual effects). While distinguishing Manski's endogenous and contextual effects would clearly be of interest, we consider solving the self-selection issue as a precondition, and leave the identification of particular mechanisms for future research.

## 3.    Some methodological issues

Despite an increasing interest, there is still no consensus regarding the magnitude and even the existence of neighbourhood effects in previous empirical work (Jencks and Mayer, 1990; Dietz, 2002). This may be due to the great diversity regarding the data, methods and variables used to test for the existence of neighbourhood effects, in particular regarding the way researchers correct or not for the endogeneity of residential locations (Dietz, 2002). Table 1 illustrates this lack of consensus by comparing results obtained by ten selected papers focusing on the potential impact of neighbourhood effects on labour-market outcomes. These papers have been chosen to reflect pioneering works that do not control for endogeneity (the first four studies) as well as more recent studies proposing various ways to deal with this problem. Among earlier works, Datcher (1982) and Case and Katz (1991) both find evidence of neighbourhood effects on the labour-force participation of young men. Osterman (1991) also finds convincing support for the existence of neighbourhood effects on the welfare participation of single mothers. However, Corcoran et al (1992), extending Datcher's pioneering study, find no evidence of neighbourhood effects and attribute this contradictory result to the wider range of individual controls used. Among studies trying to deal with the endogeneity issue, three find evidence of neighbourhood effects (O'Regan and Quigley, 1996; Cutler and Glaeser, 1997; Fieldhouse and Gould, 1998). However, Katz et al (2001), using quasi-experimental data from a government housing relocation program, find no impact of residential changes on adults' labour-market outcomes. Finally, Plotnick and Hoffman (1999) and Weinberg et al (2004) both find evidence of neighbourhood effects when endogeneity is not accounted for, but much smaller or non-existent effects when endogeneity is accounted for.

Before moving to the formal exposition of the endogeneity issue and to the enumeration of solutions that have been proposed, it is useful to briefly review two of the other methodological concerns raised when one intends to initiate an empirical evaluation of neighbourhood effects: the choice of spatial scale and the choice of measures to characterize neighbourhoods (see Dujardin, 2006, for more details on these methodological issues).

| Study | Outcome [1] and population | Context [2] | Neighbourhood: definition and characteristics | Method | Main findings |
|---|---|---|---|---|---|
| Datcher (1982) | Earnings (education); young adult men | SMSAs | Zip-code areas; average income, racial composition | OLS regressions | Neighbourhood variables are significant |
| Case and Katz (1991) | Labour-market activity (several other outcomes); young adult men | Boston | Street blocks; average outcome measures | Probit models | Strong evidence of peer effects |
| Osterman (1991) | Welfare participation; single mothers | Boston | Zip-code areas; zip-code dummies | Logit models | Neighbourhood effects are powerful; 5 neighbourhoods exert a significant effect |
| Corcoran et al (1992) | Earnings, hours of work; young male household heads | SMSAs | Zip-code areas; median income, unemployment rate, single mothers, public assistance | GLS regressions; wide range of individual controls | No evidence of neighbourhood effects |
| O'Regan and Quigley (1996) | Employment idleness; youth (16-19) at home | 4 New Jersey cities | Census tracts; adult employment, racial composition, public assistance, job access | Logistic regression; sample restricted to youth living with parents | Neighbourhood variables have a clear effect |
| Cutler and Glaeser (1997) | Idleness, earnings (education); young people (20-30) | MSAs | City-level segregation indices | OLS and instrumental variables estimates; inter-city scale | Blacks are significantly worse-off in more segregated cities |
| Fieldhouse and Gould (1998) | Unemployment | Great Britain | Travel-to-Work areas; unemployment rate, social and racial composition | Multilevel models | Geographical variables are important |
| Plotnick and Hoffman (1999) | Income-needs ratio (pregnancy, schooling); sisters aged 17-24 | SMSAs | Census tracts; income, single mothers, public assistance | Siblings data and family fixed-effects models | Evidence of neighbourhood effects in OLS but not in fixed-effect models |
| Katz et al (2001) | Economic self-sufficiency (child well-being); MTO participants | Boston | Census tracts; poverty rate, welfare receipt, single mothers | Comparison of outcomes between Experimental, Section 8 and Control groups | No impact of vouchers on employment, earnings or welfare participation |
| Weinberg et al (2004) | Annual hours worked; young adult men | MSAs | Census tracts; adult employment rate, job access | Panel regressions; individual fixed-effect models | Neighbourhood effects are present but naïve estimates are overestimated by a factor 2 to 5 |

[1] This table is limited to neighbourhood effects on labour-market outcomes. If the cited paper studies in addition other types of outcomes, these are indicated in brackets but findings are only reported for labour-market outcomes. [2] (S)MSAs: (Standard) Metropolitan Statistical Areas.

Table 1: Comparison of 10 selected studies on neighbourhood effects on labour-market outcomes

### 3.1. Some scale concerns

The first question one has to answer when initiating a research project on neighbourhood effects is the choice of a database on which to conduct analyses. In this respect, two important concerns are the level of data aggregation and the definition of neighbourhood size.

Addressing the issue of data aggregation first, empirical studies can be divided into two groups according to whether they use individual-level data or spatially-aggregated data (Sampson et al, 2002). In the first group, studies generally regress an individual outcome measure (for example an individual's employment status) on individual and family characteristics (age, gender, level of education, etc) and on a set of variables describing the social composition of the neighbourhood (for example, mean educational level). In the second group, researchers focus on data at a spatially-aggregated level (for example, census tracts) and try to explain spatial variations of the outcome of interest (mean unemployment rate) using indicators of the local social composition. The main problem with such an approach is the risk of ecological fallacy, an interpretation error arising when one tries to infer individual-level relationships from results obtained at an areal level of aggregation (Wrigley, 1995). Robinson (1950) was the first to provide empirical evidence for this potential problem. He showed that literacy and foreign birth in the US were positively associated at the State level (suggesting that foreigners were more likely to be literate than the native-born). In contrast, at the individual-level the correlation was negative. Moreover, in the context of neighbourhood effects, studies that use spatially-aggregated data are unable to distinguish neighbourhood effects from simple compositional effects (Oakes, 2004). Indeed, a correlation between an area's unemployment rate and its mean level of education can either be due to the fact that the unskilled are more often unemployed or the fact that the spatial concentration of unskilled generates poor social networks. Dujardin et al (2004) provide evidence of the superiority of individual level data in neighbourhood effect studies.

A second and related scale concern is the choice of appropriate delineations for neighbourhoods. Due to data availability, virtually all studies rely on geographic boundaries defined administratively (such as census tracts or zip-code areas) and assume that two individuals living in the same administrative unit have more contacts than individuals living in different units. While it is unclear whether these administrative boundaries accurately represent the neighbourhood conditions that really make a difference in people's lives (Ellen and Turner, 1997), due to confidentially restrictions, researchers have often no other choice but to use such artificial neighbourhood definitions.

### 3.2. How to characterize neighbourhoods

The second issue that arises is the choice of a set of indicators which may be used to measure and characterize the social composition of neighbourhoods. In this respect, several tendencies are observed. First, some authors do not characterize the social composition of neighbourhoods *per se* but use several dummy variables to reflect the individual's location in a particular area/neighbourhood (see for example, Osterman, 1991 which uses 16 dummies to reflect different areas in Boston). This method seems quite simple but becomes relatively untreatable as soon as the number of areas increases. More often, researchers use one or several quantitative measures of the social composition of neighbourhoods (for example, average family income, unemployment rate, racial composition, percentage of single-mother households or poverty rates; see column 4 in Table 1). However, while it is likely that individual outcomes are influenced by a wide variety of neighbourhood characteristics, introducing all of them into a regression analysis may cause collinearity problems as many indicators of neighbourhoods' social composition are

highly correlated (Johnston et al, 2004). The symptoms of such collinearity problems are generally instability in parameter values and significance levels (O'Regan and Quigley, 1996, present an illustration of this parameter instability). For this reason, some authors prefer to use a composite measure of the social composition of neighbourhoods. For example, Buck (2001) characterizes a neighbourhood's level of social exclusion by using a composite indicator of housing tenure, housing density, unemployment and car ownership. Similarly, Duncan and Aber (1997) use a factorial analysis to summarize thirty-four socio-demographic variables into a small number of composite factors. Dujardin et al (2008) also use factorial analyses combined with clustering techniques to define five types of neighbourhoods and use these categories as indicators of different socioeconomic environments.

## 4. The endogeneity of residential locations

The most difficult methodological problem in the study of neighbourhood effects is probably the endogeneity of residential choices which may lead the researcher to confound neighbourhood effects with simple correlated effects. The objective of the following subsections is first to provide a precise formulation of the problem and then to review the various solutions that have been proposed –sometimes wrongly– in order to try to solve it.

### 4.1. Definition

Endogeneity arises because residential locations are not exogenously determined (Dietz, 2002). On the contrary, individuals/households have some degree of choice regarding the neighbourhoods in which they live. Therefore, individuals with similar socioeconomic characteristics, notably similar labour-market outcomes, tend to sort themselves in certain areas across urban space. For example, individuals with well-paid jobs will choose to reside in better-off neighbourhoods in order to benefit from a social environment of better quality. There is thus a two-way causality: on the one hand, individual outcomes influence the choice of a residential location, while, on the other hand, residential location influences in turn individual socioeconomic outcomes. Of course, standard approaches to the estimation of neighbourhood effects allow one to control for some individual and household characteristics that might influence both neighbourhood choices and individual outcomes, as illustrated by the $X_i$ in the following equation (the same notation as in equation 1 is used but, for the sake of simplicity, Manski's endogenous and contextual effects are not distinguished anymore; $N_{n(i)}$ is used instead for indicating neighbourhood characteristics more generally):

$$y_i = \alpha + \beta' X_i + \gamma' N_{n(i)} + \varepsilon_i \qquad [3]$$

However, it is likely that some individual and household characteristics are in fact unobserved (and therefore not included in $X_i$) and influence both the outcome of interest $y_i$ (for example labour-market outcome) and neighbourhood choice $N_{n(i)}$. For example, individuals with a low labour-market attachment or with low abilities (which decreases their labour-market performance) may choose to reside in deprived neighbourhoods for economic or social reasons. As a consequence, what the researcher perceives as a neighbourhood effect through the estimated $\gamma$ parameter may simply be a spurious correlation reflecting common residential choice (Weinberg et al, 2004). Such unobserved individual/household characteristics are in fact incorporated in the error term $\varepsilon_i$, thus generating a correlation between $\varepsilon_i$ and the observed

regressor $N_{n(i)}$ (recall that in equation 2, this error term had been decomposed into an individual specific error term $\varepsilon_i$ and a group specific error term $v_{n(i)}$; equation 3 now assumes that $v_{n(i)}$ being unobserved, it cannot be distinguished from $\varepsilon_i$). Therefore, the consistency assumption in OLS estimation methods stating that regressors must be uncorrelated with the error term is not valid as

$$E\left(\varepsilon_i \middle| X_i, N_{n(i)}\right) \neq 0 \qquad\qquad [4]$$

and results of studies based on standard methods that do not control for these unobserved characteristics will be biased (a formulation of the bias can be found in Greene, 2008). As it arises from the presence of unobserved individual and household characteristics, the endogeneity bias might also be understood as an omitted variable bias and a consistent estimation of equation 3 would require constructing a consistent estimate of $E(\varepsilon_i/X_i, N_{n(i)})$ and using it as an additional regressor (Durlauf, 2004).

The direction of the bias is difficult to predict as it depends on the relationship between the unobserved factors that determine neighbourhood choices and the unobserved factors that determine the outcome of interest (Evans et al, 1992). However, researchers generally assume that neighbourhood effects are *overestimated*. For example, in the context of neighbourhood effects on children outcomes, parents that lack the financial resources to move to better neighbourhoods often lack the qualities to help their children to perform well at school; then the true $\gamma$ parameter will be overestimated leading to an *upward* bias. However, a *downward bias* can also arise if parents choose a single-earner strategy and earn less, therefore being forced to live in poor neighbourhoods, but allowing the parent who stay at home to spend more time with their children (Duncan et al, 1997). The following subsections review the methods that have been proposed in order to cope with this issue. The first method is based on experimental data, where exogeneity is reached through random assignment of individuals in different neighbourhoods, whereas the following methods use standard observational data collected through conventional surveys (such as census data).

## 4.2. Quasi-experiments

Perhaps the best way to correctly identify the causal effects of neighbourhoods would be to realize a kind of controlled experiment in which individuals would randomly be assigned to neighbourhoods (Durlauf, 2004). One could then compare the socioeconomic outcomes of similar individuals with respect to the characteristics of the neighbourhood in which they are assigned. While it is difficult to translate such controlled experiments to the study of neighbourhood effects, data provided by government subsidized relocation programs in the United States can be considered as *quasi-experiments* (see Oreopoulos, 2003, for a review of such programs). Indeed, government interventions into the residential choices of households can be used to assess neighbourhood effects, as households that would normally belong to one neighbourhood are moved to another through an exogenous intervention. The best-known example is the *Moving To Opportunity* Program conducted by the Department of Housing and Urban Development in five American cities (Baltimore, Boston, Chicago, Los Angeles and New York). In this program, eligible families (public housing families with children residing in census tracts with at least 40% of poverty) who volunteered for the program were randomly drawn from the waiting list and randomly assigned into three groups: the Experimental group, in which families received housing subsidies and search assistance to move to private-market housings in low-poverty census tracts, the Comparison group, in which families received geographically unrestricted housing subsidies but no search assistance, and the Control group,

which received no special assistance. Households were periodically surveyed in years following relocation, thus allowing researchers to analyse the resulting evolution of their socioeconomic characteristics (see for example Katz et al, 2001 and Kling et al, 2007).

While it is clear that such data are well-suited for removing the endogeneity bias in neighbourhood effects studies, it does present some disadvantages. The first issue is undoubtedly the cost and difficulty of implementation as well as the ethical concerns surrounding such experiments (Harding, 2003). Moreover, eligibility conditions, screening of families and participation willingness generate sample selection problems (Dietz, 2002; Durlauf, 2004). Indeed, the estimates give the effects of the relocation program on the types of people who were authorized and choose to participate, and may thus be different from the effects obtained from relocating a randomly selected group of families from poor areas. Brock and Durlauf (2001) also suggest that the implementation of such programs to a wider scale would result in moves of large numbers of households, thus modifying the social composition of neighbourhoods. Therefore, estimation of neighbourhood effects without taking into account induced changes in the neighbourhood composition would be misleading and would not allow the generalization of results to wide-scale programs.

### 4.3. Sample restriction

The second method is perhaps the easiest to implement and the most frequently used. It consists of restricting the studied sample to a group of individuals for whom residential choices are limited and might thus be considered as fairly exogenous. For example, in their study of neighbourhood effects on the employment outcomes of young adults, O'Regan and Quigley (1996) limit themselves to a sample of youngsters still living with at least one parent. They justify their approach by the fact that location choice has been made previously by the parents and can thus be thought of as exogenous to the employment status of their children.

However, the sample restriction method presents some disadvantages (Glaeser, 1996; Ihlanfeldt and Sjoquist, 1998). First, it is only applicable to very limited sub-populations (generally the under-25) and one cannot use it when the interest is on elderly adults, nor can results be generalized to young adults that have moved out of parental home. Moreover, it can create a sample selection bias. Indeed, in the case of neighbourhood effects on labour-market outcomes, young adults obtaining a job are more likely to leave parents' home than young adults still unemployed, and this leaving rate might differ according to the perceived characteristics of neighbourhoods. Finally, it does not completely eliminate endogeneity bias. Indeed, the household's residential choice depends on observed as well as unobserved parental characteristics and some of these unobserved parental characteristics might also influence children employment outcomes. For example, lack of commitment to work or social norms may induce parents to locate in high poverty neighbourhoods and also probably influence youngster's motivation and intensity of job search (Glaeser, 1996).

### 4.4. Siblings data and family fixed effects

Some studies resort to siblings data to solve endogeneity, i.e. data on individual children from the same family (Aaronson, 1998; Plotnick and Hoffman, 1999). It consists of finding siblings pairs in which one sibling was exposed to a particular neighbourhood and the other one to another neighbourhood (because of family residential moves). For each sibling $i$=1, 2 belonging to family $f$, the outcome of interest ($y_{if}$) is

then expressed as:

$$y_{if} = \alpha' X_f + \beta' X_{if} + \gamma' N_{n(i)} + \varepsilon_f + \varepsilon_{if} \qquad [5]$$

where $X_f$ is a vector of family-specific variables, $X_{if}$ is a vector of individual characteristics for sibling $i$ belonging to family $f$, $N_{n(i)}$ is the standard vector of neighbourhood characteristics, $\varepsilon_{if}$ is an error term associated with sibling $i$ (individual unobserved characteristics) and $\varepsilon_f$ is an error term associated with family $f$ (family unobserved characteristics). Then, by assuming that observed and unobserved family characteristics are constant across time (and thus for the two siblings), differencing outcomes between siblings eliminates family effects and $\gamma$ is interpreted as an unbiased estimate of neighbourhood effects:

$$y_{1f} - y_{2f} = \beta'(X_{1f} - X_{2f}) + \gamma'(N_{n(1)} - N_{n(2)}) + (\varepsilon_{1f} - \varepsilon_{2f}) \qquad [6]$$

The application of this method is quite limited due to data availability as it requires data on families that have raised two or more children in different neighbourhood conditions. Moreover, the assumption on which it rests is quite strong as it requires that parents do not change their behaviour when moving (Durlauf, 2004). However, residential moves might reflect changes in parents' unobserved characteristics that may also influence children outcomes (change in income, divorce). Therefore, such estimates of neighbourhood effects might in fact reflect changes in family unobservables.

## 4.5. Longitudinal data and individual fixed effects

Recent studies have used longitudinal data that allows researchers to track individuals over time and follow their successive residential moves (see for example Weinberg et al, 2004). By comparing individual outcomes before ($t=1$) and after the move ($t=2$), one can assess neighbourhood effects using the equations below:

$$y_{it} = \alpha + \beta' X_{it} + \gamma' N_{n(it)} + \varepsilon_{it} \qquad [7]$$

$$y_{i2} - y_{i1} = \beta'(X_{i2} - X_{i1}) + \gamma'(N_{n(i2)} - N_{n(i1)}) + (\varepsilon_{i2} - \varepsilon_{i1}) \qquad [8]$$

$X_{it}$ is a vector of individual characteristics for individual $i$ at time $t$, some of these being constant across time (such as gender) while others may vary across time (such as the level of education or the marital status). $N_{n(it)}$ is a vector of characteristics of the neighbourhood in which the individual $i$ resides at time $t$. However, this method suffers from the same problems as siblings studies. Data availability concerns limit its usefulness and it relies on the assumption that unobserved individual characteristics ($\varepsilon_{it}$) that influence socioeconomic outcomes are constant across time. Yet, it is likely that residential moves are the consequence of previous changes in individual characteristics, including observables such as the family composition as well as unobservable determinants.

## 4.6. Multilevel models

Multilevel models (also known as hierarchical, mixed or random-coefficient models) have also been used to study the influence of neighbourhoods on individual outcomes, particularly in the context of health studies (Blakely and Subramanian, 2006). Without going into details, the principle of multilevel modelling is that the parameters of the model are not constant but are allowed to vary across neighbourhoods:

$$y_i = \alpha_{n(i)} + \beta_{n(i)}' X_i + \varepsilon_i \qquad \text{with} \qquad \alpha_{n(i)} = \alpha + \upsilon_{n(i)} \qquad \text{and} \qquad \beta_{n(i)} = \beta + \mu_{n(i)} \qquad [9]$$

In this formulation, $\alpha_{n(i)}$ and $\beta_{n(i)}$ are random variables, with means $\alpha$ and $\beta$. $\upsilon_{n(i)}$ and $\mu_{n(i)}$ are neighbourhood's deviations from mean values and reflect the fact that the relationship between the outcome and the individual-level variables varies across neighbourhoods. Note that these random coefficients can further be explained by neighbourhood-level explanatory variables (see Goldstein, 2003 for more detailed formulations).

In other contexts, multilevel models present numerous advantages. Notably, they allow one to take into account the fact that, because of the grouping of individuals within neighbourhoods, individuals from the same neighbourhood are generally more similar than individuals drawn randomly from the population, which violates the assumption of standard OLS regressions that individual observations must be independent (Goldstein, 2003). However, in the context of epidemiologic studies, it has been argued that, by explicitly considering the grouping of individuals into neighbourhoods, multilevel models allow one to obtain unbiased estimates of neighbourhood effects and to perfectly distinguish neighbourhood effects from the effects of individual characteristics (which are often termed "compositional" effects in this context). As is argued by Oakes (2004), this assertion is a little bit presumptuous. Indeed, as they do not take into account the reciprocal nature of the relationship between neighbourhood and outcome, multilevel models are unable to specify whether the estimated spatial variations of the relationship between $y_i$ and $X_i$ are the result of neighbourhood effects or the result of non-random neighbourhood selection (i.e. correlated effects). Indeed, spatial variations in $\alpha_{n(i)}$ and $\beta_{n(i)}$ may simply indicate that some unobserved individual characteristics have an impact on the outcome and simultaneously determine neighbourhood choice.

## 4.7. Inter-city studies

A sixth group of studies have used inter-city data in order to evaluate the effects of residential segregation on individual outcomes, and to achieve exogeneity. The best known example of such study is Cutler and Glaeser (1997). In their study of the effect of racial and income segregation on individual outcomes, these authors argue that the selection bias arising from the fact that more successful African-Americans will choose to live in richer and whiter neighbourhoods is difficult to solve using an intra-urban approach. Instead, by focusing on inter-city data, it is possible to evaluate if African-Americans in more segregated cities on average have worse or better outcomes than African-Americans in less segregated cities. By doing this, one avoids the within-city sorting of individuals along abilities and other unobserved characteristics.

Such studies however present some shortcomings. First, one can question the comparability of segregation measures across different metropolitan areas. Indeed, the effect of the size of spatial units on the value taken by such indices is well documented (see for example Wong, 2004, for a discussion of the Modifiable Areal Unit Problem in the context of segregation indices). By comparing segregation indices across metropolitan areas, one implicitly assumes that spatial units in different cities are of comparable size and shape. This seems a rather strong assumption. Moreover, the constructed segregation index applies to all people residing in the same metropolitan area, regardless of their specific location within the city. Yet, there can be substantive variations in social composition within the city itself (Ihlanfeldt and Sjoquist, 1998). In the extreme case where every metropolitan area has the same degree of residential segregation

but where important disparities exist between neighbourhoods in the same city, such an analysis would see no segregation effects at all. A huge cost in terms of loss of information is thus attached to such strategies (Glaeser, 1996).

## 4.8. Instrumental variables

More generally, in econometrics, endogeneity is dealt with by using instrumental variable techniques. Broadly speaking, this method consists of replacing the endogenous regressor by an instrument, i.e. a variable which is highly correlated with the endogenous regressor but not with the unobserved determinants of the outcome. This therefore eliminates the correlation between the regressor and the error term (see Greene, 2008 for an introduction to instrumental variables). The most common form of Instrumental Variable Estimation (IVE) is two-stage least squares (2SLS). In the first stage, the endogenous regressor (in the case of neighbourhood effect studies, this would be the neighbourhood characteristic $N_{n(i)}$) is regressed on a set of chosen instruments. The resulting coefficient estimates are used to generate a predicted value for the neighbourhood characteristic. In the second stage, this predicted value is used in the outcome equation, in place of the actual neighbourhood characteristic. If the instruments are correctly specified (i.e. they are highly correlated with the neighbourhood characteristics but are not direct determinants of the outcome), then the predicted value from the first-stage equation is uncorrelated with the error term of the outcome equation.

Instrumental variables are rather common in econometric studies. However, they have rarely been applied to the study of neighbourhood effects. This is mainly because it is hard to find a good instrument in this particular context (Dietz, 2002). Indeed, one has to find a variable that is a true determinant of neighbourhood choice but that is not correlated with the outcome measure. One can easily figure out how tricky it is to find such an instrument, and the few examples of instrumental variable estimates of neighbourhood effects have received some criticism (Durlauf, 2004). For example, in their study of neighbourhood influences on educational outcomes, Duncan et al (1997) use the characteristics of the mother's neighbourhood of residence after the child has left the parental home. This assumption relies on the fact that as long as the child is at home, the neighbourhood of residence reflects both mother's preferences as well as her (unobserved) concerns about the effects of neighbourhood on child's development. After the child has left home, mother's neighbourhood choice no longer reflects these concerns but only her own preferences. However, the authors themselves criticize their instrument choice on the basis that inertia may cause mother to stay in the same neighbourhood even after children have left home. Evans et al (1992) also use IVs to study the effect of peer groups on teenage pregnancy and school dropout. They instrument the percentage of disadvantaged pupils in the school by measures of unemployment and poverty rates at the metropolitan level, based on the assumption that adolescents living in a metropolitan area with a high poverty rate are more likely to attend a school with a higher percentage of disadvantaged students. This choice of instrument is debatable as it implies exogeneity in parent's choice of a metropolitan area and it is not clear how such instruments can account for neighbourhood effects within cities (Durlauf, 2004).

While the validity of the instruments used by Evans et al (1992) may be questioned, their results are quite interesting as these authors provide peer group effects estimates obtained with and without treatment for endogeneity. First, they regress the propensity for a teenage girl to become pregnant on a set of family characteristics as well as on the log of the percentage of disadvantaged students in her school. They find a positive and significant coefficient for this last variable, thus indicating the presence of peer group effects.

Then, using an instrumental variable specification, they find no evidence at all for the existence of peer group effects, concluding that the significant effect in the first specification could in fact be attributed to unobserved family determinants affecting the choice of a school for their girl as well as the girl's propensity of becoming pregnant. Their findings thus strongly advocate in favour of an explicit treatment of endogeneity in all studies of peer group and neighbourhood influences on individual outcomes.

## 4.9. Conclusion

While initial studies of neighbourhood effects rarely mentioned the problems associated with endogeneity, nearly all recent empirical studies are aware of its existence and attempts to solve it are numerous. The results of comparative studies suggest that the biases arising from not taking the endogeneity of residential location choices into account are important (Evans et al, 1992; Plotnick and Hoffman, 1999; Weinberg et al, 2004). Unfortunately, the perfect solution does not exist and the various methods reviewed in this paper either rely on very particular datasets that are rarely available (for example experimental data or siblings data) or present several important shortcomings. Durlauf (2004) suggests that future empirical works should attempt to simultaneously model neighbourhood effects and neighbourhood configurations via structural models: *"Structural models will allow for a full exploration of self-selection in neighbourhoods models"* (Durlauf, 2004, p. 2232). To our knowledge, there have been only few attempts of such structural modelling in the context of neighbourhood effects and this will have to be on the research agenda for the next years.

## 5. Empirical application: labour-market outcomes of young adults in Brussels

The remainder of this paper is devoted to an empirical illustration of the study of neighbourhood effects and how to cope with endogeneity issues. It presents results of an empirical exercise aiming at estimating neighbourhood effects on labour-market outcomes of young adults residing in the Brussels urban area. Testing all the methods presented in the previous section to cope with endogeneity would be an impossible task as most of these methods rely on particular datasets that are not available for Brussels. Instead, the endogeneity of residential choices will be treated with the sample restriction method, i.e. restricting the sample to young adults residing with their parents. In addition, a step-by-step model-building strategy will be used in order to provide an assessment of the sensitivity of the results in the presence of observed and unobserved parental covariates. This allows evaluating the robustness of the results. This method had already been applied by the authors on 1991 Census data in a previously published paper (Dujardin et al, 2008). The analyses presented here provide an update of the results of this previous paper with 2001 census data. The following subsections proceed by briefly describing our research question (i.e. why residential location should influence labour-market outcomes), the study area and the data used. Then, the methods and results will be presented in details.

### 5.1. Research question

Estimates of neighbourhood effects on individual outcomes are here illustrated in the particular context of labour-market outcomes. There are numerous reasons why residential locations might influence labour-market outcomes. First, the socioeconomic composition of the neighbourhood can influence *human capital acquisition*, especially for young adults, which may in turn deteriorate their employability in later years. Indeed, as the success of a given student depends on the results of other students in his class, in neighbourhoods with a high concentration of low-ability students, peer effects can deteriorate school

achievements and employability (Benabou, 1993). Social problems which deteriorate employability (like dropping out of school, consuming drugs or having illicit activities) also spread through social interactions within the neighbourhood (Crane, 1991). This contagion is all the more prevalent as adults are themselves unemployed and do not provide a role model to which youngsters could identify (Wilson, 1987). Moreover, the socioeconomic composition of the neighbourhood influences the quality of *social networks*. This is a crucial point since a significant portion of jobs are usually found through personal contacts and since low-skilled workers, young adults, and ethnic minorities often resort to such informal search methods (Holzer, 1988). Therefore, in neighbourhoods where local unemployment rates are higher than average, local residents know fewer employed workers that could refer them to their own employer or provide them with professional contacts. Finally, employers may be reluctant to hire workers residing in disadvantaged neighbourhoods (a practice known as *territorial discrimination*; Zenou and Boccard, 2000). Note that the spatial mismatch hypothesis, which emphasizes the role of the physical disconnection between job opportunities and residential locations, has also been put forward to explain the poor labour-market outcomes in the context of black ghettos in American cities (see Gobillon et al, 2007 for a survey). Previous work has suggested this hypothesis has no support in Brussels; we therefore will not elaborate further on this issue (Dujardin et al, 2008).

## 5.2. Data and studied area

### 5.2.1. Studied area and neighbourhood size

In institutional terms, Brussels comprises 19 municipalities and hosts around 1 million inhabitants on a 163 km² area. However, as in most cities, Brussels' functional metropolitan area extends far beyond its institutional limits. Therefore, the so-called Extended Urban Area is used to reflect its functional metropolitan area, which comprises 41 municipalities that host together 1.4 million inhabitants and extends on a 723 km² area.

The smallest spatial unit for which census data are officially available is the statistical ward (in French, "secteur statistique"), a subdivision of the municipality defined according to social, economic and architectural similarities (Brulard and Van der Haegen, 1972). Statistical wards that present a common functional or structural nature (for example a common attraction pole like a school or a church) can further be grouped into larger entities, which constitute an intermediate level between the statistical ward and the municipality. The limits of these units were generally defined by geographical obstacles, like important roads, railways or waterways. Because of this delineation criterion and their functional definition, these units can be considered as appropriate to reflect social influences and are used here to define the local environment that may potentially matter for individuals. There are 328 such neighbourhoods in the Extended Urban Area, grouping on average 4,250 inhabitants. For statistical reasons, neighbourhoods with less than 200 inhabitants are not considered in this analysis (i.e. 19 neighbourhoods were left aside).

### 5.2.2. Data and neighbourhood characteristics

The statistical analyses are based on data extracted from the 2001 Socioeconomic Survey carried out by the Belgian National Institute of Statistics, which provides data for all individuals residing in Belgium (i.e. this is a 100% census). Analyses were restricted to members of private households. For each individual, detailed information on personal characteristics (including age, gender, education, citizenship, employment status, kinship with household's head) are provided, along with family and housing characteristics (for instance, type of family or car ownership) as well as statistical ward (and

neighbourhood) of residence.

These statistical analyses aim at explaining an individual's unemployment probability by taking into account personal and household characteristics as well as the possible role played by the neighbourhood of residence. In this purpose, it is necessary to choose one or several measures of the socioeconomic composition of neighbourhoods. As mentioned previously, even though it is likely that individual outcomes are determined by a wide variety of neighbourhood characteristics (such as mean income of households in the neighbourhood, unemployment rate or racial composition), considering all these together into a single regression may cause collinearity problems (Johnston et al, 2004). Therefore, a typology of neighbourhoods is used, which is intended to reflect different types of social environments within Brussels. This typology is built on a set of eleven neighbourhood characteristics, chosen in order to reflect various aspects of the social composition of neighbourhoods likely to affect labour-market outcomes. These variables concern educational levels, professional statuses, unemployment rates, percentages of foreigners, of single-mother households as well as average household income. First, a Principal Component Analysis is run in order to define a limited number of non-correlated factors summarizing the information carried by this set of neighbourhood variables. Then, neighbourhoods are grouped according to their coordinates on the factorial axes, using a hierarchical ascending classification (with the Ward method which minimizes intra-group variance), in order to define deprived neighbourhoods versus not deprived neighbourhoods (in a previous paper, a 5 classes typology was used; the two classes typology is used here for the sake of simplicity; see Dujardin et al, 2008 for more details on this classification).[4]

### 5.2.3. *Spatial structure of Brussels*

The Brussels Extended Urban Area presents a well-marked spatial structure characterized by important disparities opposing its city centre to the periphery. Figure 1 maps the percentage of unemployed workers among labour-force participants aged 19 to 64 in 2001, highlighting a zone of very high unemployment rates (above 20%, and even above 30% for some neighbourhoods) in the central part of the urban area, along the former industrial corridor. On the contrary, unemployment is much lower (below 12.5% or even 8.5%) in the suburbs. Figure 2 maps deprived neighbourhoods and table 2 summarizes their characteristics. Deprived neighbourhoods are located in the centre of Brussels. They are characterized by high unemployment rates (2.5 as high as the average unemployment rate in not deprived neighbourhoods), high proportions of North-Africans and Turks and single-mother households, low educational levels, high percentage of blue-collars, and low income levels.

---

[4] The neighbourhood typology used here was defined by Dujardin et al (2008) on the basis of 1991 Census data. This 1991 typology is used here to explain individual unemployment propensities in 2001. Because of inertia in housing prices and residential choices, it is likely that similar results would have been obtained for a typology of neighbourhoods on the basis of 2001 data.
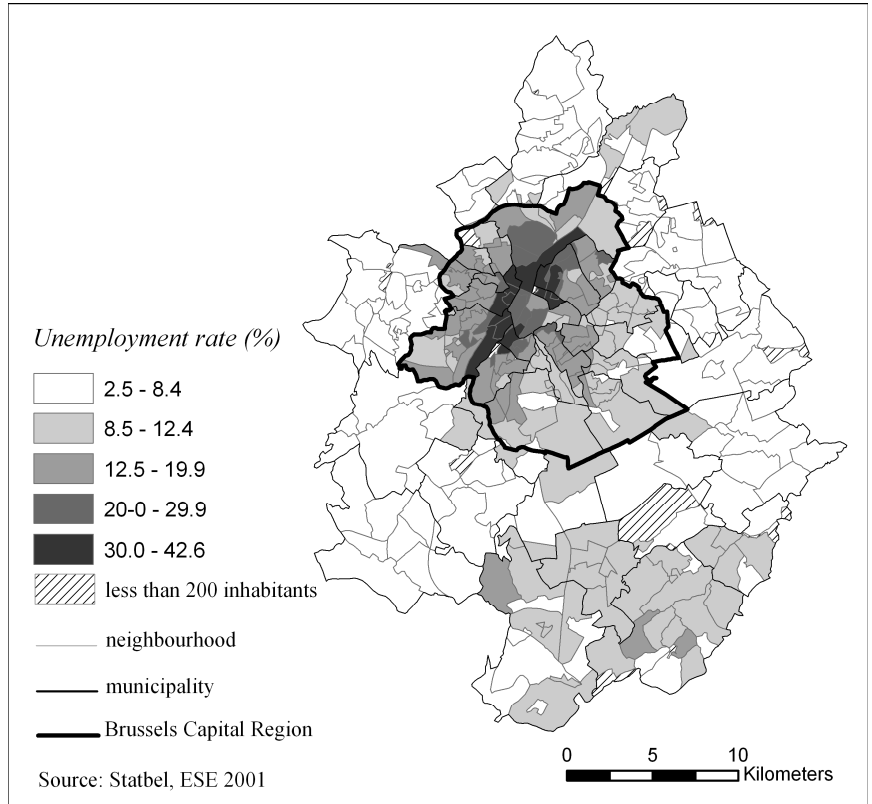
Figure 1: Percentage of unemployed workers among labour-force participants in the Brussels E.U.A. in 2001
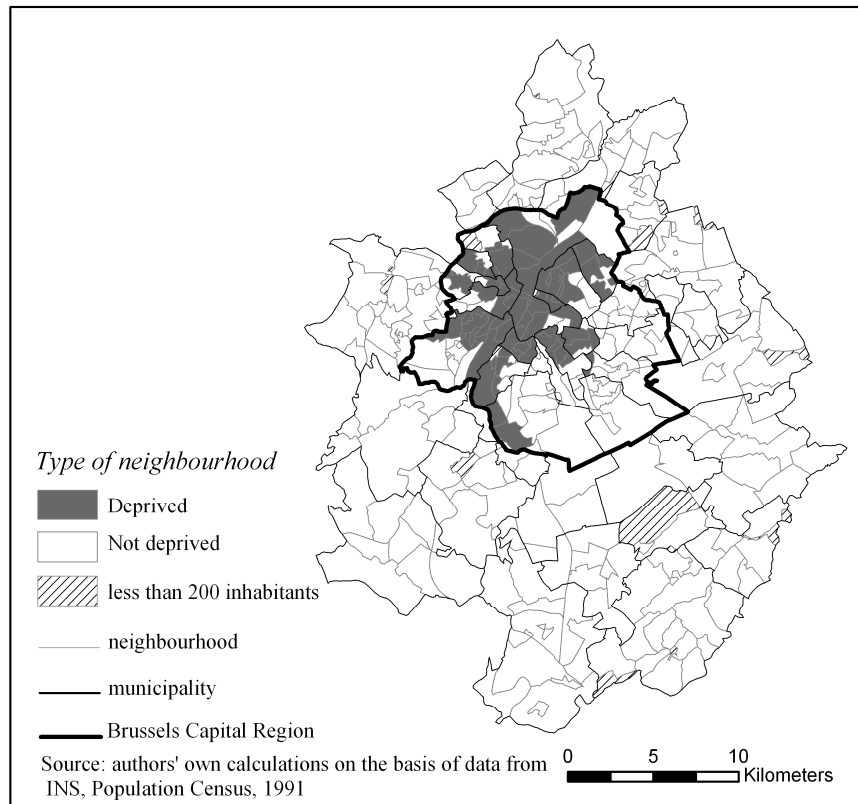
Figure 2: Location of deprived neighbourhoods in the Brussels E.U.A.

|                                          | **Deprived** | **Not deprived** | **Total** |
|------------------------------------------|--------------|------------------|-----------|
| Demography                               |              |                  |           |
|    % North-Africans and Turks | 16.1     | 0.7              | 4.3       |
|    % single-mother households | 23.5     | 15.8             | 17.6      |
| Average household income                 | 772          | 1,113            | 1,033     |
| Education                                |              |                  |           |
|    % of students in technical classes | 39.7 | 25.5         | 28.8      |
|    % with lower education  | 53.9         | 43.2             | 45.7      |
|    % with at least intermediate education | 28.6 | 46.2     | 42.1      |
|    % with higher educational levels | 13.0 | 23.8             | 21.3      |
| Professional status                      |              |                  |           |
|    % blue-collars          | 31.7         | 16.2             | 19.8      |
|    % executives            | 7.4          | 14.0             | 12.4      |
| Unemployment                             |              |                  |           |
|    Unemployment rate (19-64) | 18.8       | 7.4              | 10.0      |
|    Youth unemployment rate (19-25) | 26.5  | 15.0             | 17.7      |
| Total population                         | 563,704      | 829,106          | 1,392,810 |
| Number of neighbourhoods                 | 72           | 237              | 309       |

Source: calculations based on data from the 1991 Population Census (INS)

Table 2: Mean characteristics of neighbourhood types

19

## 5.3. A logistic model of unemployment probability with sample restriction

Individual-level data are used to estimate a logistic model of unemployment probability while taking into account both personal and household characteristics as well as the neighbourhood of residence, as illustrated in the following equation:

$$Log\left(\frac{P_i}{1-P_i}\right) = \alpha + \beta_1'I_i + \beta_2'H_i + \gamma DN_i \qquad [10]$$

where $P_i$ is the unemployment probability of individual $i$, $I_i$ is a vector of personal characteristics, $H_i$ is a vector of household characteristics and $DN_i$ is a dummy variable indicating whether the neighbourhood in which individual $i$ resides is deprived ($DN_i$=1) or not ($DN_i$=0). $\alpha$, $\beta_1$, $\beta_2$ and $\gamma$ are vectors of parameters that will be estimated using Maximum Likelihood Estimation (MLE). In particular, $\gamma$, when significantly different from zero identifies an impact of neighbourhood deprivation on unemployment, i.e. *neighbourhood effects*. Using [10], the individual probability of unemployment $P_i$ is given by:

$$P_i = \frac{e^{(\alpha + \beta_1'I_i + \beta_2'H_i + \gamma DN_i)}}{1 + e^{(\alpha + \beta_1'I_i + \beta_2'H_i + \gamma DN_i)}} \qquad [11]$$

In this equation, the parameter $\gamma$ is potentially subject to an endogeneity bias. As already explained in the first part of this paper, residential locations are partly determined by labour-market outcomes as individuals choose where they live and sort themselves in the urban space along their socioeconomic characteristics. In other words, the right-hand side variable $DN_i$ in equation 11 is an endogenous regressor, which is determined jointly with the left-hand side variable $P_i$, and $\gamma$ cannot be estimated without any bias using standard methods. The first part of this paper discussed various strategies used in the literature to correct for the endogeneity of neighbourhood choice. Although it is an imperfect solution (Ihlanfeldt and Sjoquist, 1998), the sample restriction method is used here in this purpose and all statistical analyses are restricted to young labour-force participants (aged 19 to 25) residing with at least one parent (as in O'Regan and Quigley, 1996). This rests on the assumption that the choice of a residential location has been made previously by the parents and is thus fairly exogenous to the employment status of their children. In addition, focusing on at-home young adults will enable the use of parental explanatory variables as the individual census database allows identifying members of the same household. Parental characteristics are indeed important determinants of children outcomes and it is important to take these into account, when possible. Setting aside individuals for whom important personal and family characteristics are missing as well as individuals living in neighbourhoods of less than 200 inhabitants, the studied sample consists of 27,044 individuals.

In a first step, equation 11 is estimated using only individual level explanatory variables as well as neighbourhood type. Parental characteristics will be introduced in a second step, in the sensitivity analysis. The set of individual explanatory variables $I_i$ includes gender, age, level of education and citizenship. Three levels of education are distinguished: lower for individuals with at most a diploma of junior secondary education (normally corresponding to an age of 15); intermediate for those with a diploma of senior secondary education (normally aged 18); and higher for those with a higher diploma. Concerning citizenship, four main groups are defined: Belgians, foreigners from the European Union, North-African

and Turkish foreigners, and other foreigners. Furthermore, the nationality of the household head is used to approximate the concept of ethnicity, by distinguishing Belgians with Belgian parents, Belgians with EU parents, Belgians with North-African or Turkish parents, and Belgians with parents of other nationality.

Results from this model are presented in Model I of Table 3, while Model II adds a dummy variable for living in a deprived neighbourhood or not. Note that all results are presented in terms of odds ratios. This means that the reference value is one (which indicates no effect) and that a value above one indicates that the corresponding variable increases the unemployment probability, while an odds ratio between zero and one indicates that the corresponding variable decreases the unemployment probability. Results show that men or educated workers are less likely to be unemployed than women or workers with a lower education. The probability of unemployment also decreases with the age of the individual. Moreover, citizenship plays a key role: North-Africans and Turks and other foreigners are more disadvantaged than UE citizens and Belgians. This is consistent with discrimination on the labour market, but may simply reflect some differences in competence and qualification that are not taken into account by the educational level and which may differ between ethnic groups (such as experience or language fluency). Interestingly, young Belgian adults born of foreign parents are more likely to be unemployed than young Belgian adults of Belgian parents, suggesting that besides citizenship, the name or visible characteristics associated with foreign origin are a handicap on the labour market. This is consistent with both labour-market discrimination as well as social networks of lower quality for individuals of foreign parents.

Introducing neighbourhood deprivation in the regression (Model II) significantly increases the fit of the model (see the likelihood ratio and the Akaike Information Criterion). The odds ratio of the neighbourhood deprivation variable is 1.491 and is significant at a 1% level, indicating that all else being equal, residing in a deprived neighbourhood significantly increases the odds of being unemployed by nearly 1.5. This indicates that, all else being equal, young adults living in neighbourhoods characterized by the worst combination of social characteristics are more likely to be unemployed, thus confirming the importance of the social environment on labour-market outcomes (through mechanisms such as peer effects, role models or poor social networks).

As was already mentioned, the sample restriction used to solve endogeneity presents some shortcomings, the most important being that it does not completely eliminate endogeneity. Indeed, the assumption on which it rests (that residential choice is made previously by the parents and is thus exogenous to the employment status of their young adult) is questionable. Indeed, it is likely that parental characteristics determining residential choices also influence children's future employment outcomes (Glaeser, 1996). In this context, Model II does not allow to distinguish neighbourhood effects from the effect of parental characteristics on the unemployment probability of young adults living with their parents.

Instead of attempting to completely remove endogeneity, it may be useful to evaluate potential remaining bias by conducting a sensitivity analysis, in order to assess the robustness of estimated neighbourhood effects (as suggested by Glaeser, 1996, p. 62). In this purpose, a two-step strategy is used. Firstly, several models of unemployment probability are estimated, which incorporate various sets of parental characteristics. The estimated neighbourhood effects from these models are compared in order to test the robustness of the results in the presence of *observed* parental covariates. The second step of the sensitivity analysis consists of generating random variables that are correlated to a certain degree with both the unemployment probability of young adults and with parental residential choice. Introducing these random

variables in the unemployment probability model tests the sensitivity of the results to *unobserved* parental covariates.

### 5.4. Sensitivity to observed parental covariates

In their study of neighbourhood effects on children outcomes, Ginther et al (2000) argue that the disparity in past estimates of neighbourhood effects is mainly the result of differences in the set of included family characteristics. They specify a number of models that vary in the extend to which family characteristics are introduced as statistical controls and show that estimated neighbourhood effects tend to fall in value (and sometimes become insignificant) as the set of family controls becomes more complete. Building on their framework, several models of youth unemployment probability are estimated, which incorporate various sets of household characteristics, moving away from a model with no household variable towards a model including an extensive set of parental and household controls. These parental characteristics were built by assigning to each young adult the characteristics of the household head or that of the household head's spouse (when data was missing for the household head). For each young adult, the parental employment status and educational level were computed, as well as the household car ownership. Single-mother households were also identified, as these households are more frequently prone to social problems detrimental to finding a job. Table 3 includes these parental characteristics step by step in a model including only youth personal characteristics and the characteristics of their neighbourhood of residence (Models III to VI). Comparing estimated neighbourhood effects in these different models allows one to test the robustness of the results to the omission of observed parental covariates.

Regarding parental characteristics, models III to VI show that the unemployment probability of a young adult is higher when the household head (or spouse) is not participating in the labour-force or is unemployed than when he/she is employed. This effect is highly significant and is consistent with social network theories (at the household level, unemployed parents being little able to help their job-seeking children) and socialization considerations (unemployed parents failing to provide their children with an image of social success to whom they could identify; Wilson, 1987). Living in a single-mother household (Model VI) also significantly increases the likelihood that a young adult is unemployed, suggesting that these households are more frequently prone to social problems detrimental to find a job. Living in a household which does not own a car (an indirect measure of lack of financial resources) also significantly increases the unemployment probability, suggesting it is more difficult for individuals from poorer households to find a job. Surprisingly, the effect of parental educational level is not always significant, and when significant, seems counter-intuitive: all other things equal, having a parent with a higher level of education increases the unemployment propensity of young adults. This counter-intuitive effect illustrates one shortcoming of the sample restriction method used to solve endogeneity. As already mentioned, restricting the sample to young adults living with parents may create a sample selection bias in the sense that young adults obtaining a job are more likely to leave parental home than young adults still unemployed. Indeed, if having parents with a high socioeconomic status (as reflected by parental educational level) in fact increases the chances to find a well-paid job, it is likely that young adults originating from these families will move out of their parents' dwelling more rapidly. This would leave an over-representation of unemployed young adults among families with higher parental socioeconomic status. Another possible explanation is that children from more educated (and richer) families do not feel pressured to intensively search for a job in order to move out of unemployment if they get financial support from their parents. In the absence of longitudinal data in which young adults are followed after

they move out of parental home, it is not possible to distinguish between these two potential explanations.

By comparing the parameters and significance levels of the neighbourhood types across the different models, one can assess the sensitivity of the estimated neighbourhood effects to the inclusion of a more comprehensive set of parental controls. Table 3 shows that although the inclusion of parental and household characteristics significantly increases the fit of the model (see the likelihood ratios), the estimated neighbourhood effects change little (ranging from 1.491 in Model II to 1.418 in Model VI) and all parameters remain significant at a 1% level.

| Model | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|
| *Likelihood ratio* | 690.91 | 842.48 | 964.55 | 1040.05 | 1109.54 | 1130.33 |
| *Akaike Information Criterion* | 28,791 | 28,641 | 28,523 | 28,452 | 28,384 | 28,365 |
| <u>Neighbourhood Type</u> | | | | | | |
| *Deprived* | | 1.491*** | 1.415*** | 1.470*** | 1.414*** | 1.418*** |
| <u>Individual characteristics</u> | | | | | | |
| *Male* | 0.937** | 0.946* | 0.951$^{NS}$ | 0.946* | 0.949* | 0.950* |
| *Age* | 0.914*** | 0.915*** | 0.908*** | 0.905*** | 0.906*** | 0.906*** |
| *Education* | | | | | | |
| *  Lower* | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| *  Intermediate* | 0.641*** | 0.661*** | 0.674*** | 0.656*** | 0.667*** | 0.669*** |
| *  Higher* | 0.695*** | 0.747*** | 0.780*** | 0.705*** | 0.724*** | 0.729*** |
| *Citizenship* | | | | | | |
| *  Belgian (of Belgian parents)* | Ref. | Ref. | Ref. | Ref. | Ref. | Ref. |
| *  Belgian (of EU parents)* | 1.111$^{NS}$ | 1.069$^{NS}$ | 1.067$^{NS}$ | 1.110$^{NS}$ | 1.120$^{NS}$ | 1.136* |
| *  Belgian (of North-African or Turkish par.)* | 1.916*** | 1.517*** | 1.351*** | 1.424*** | 1.397*** | 1.443*** |
| *  Belgian (of parents of other citizenship)* | 1.893*** | 1.689*** | 1.574*** | 1.501** | 1.469*** | 1.491*** |
| *  EU* | 1.337*** | 1.220*** | 1.198*** | 1.246*** | 1.256*** | 1.268*** |
| *  North African and Turkish* | 3.203*** | 2.543*** | 2.267*** | 2.363*** | 2.297*** | 2.374*** |
| *  Other* | 2.815*** | 2.548*** | 2.435*** | 2.342*** | 2.159*** | 2.175*** |
| <u>Parental and household characteristics</u> | | | | | | |
| *Employment status and professional status* | | | | | | |
| *  Employed* | | | Ref. | Ref. | Ref. | Ref. |
| *  Not participating to labour force* | | | 1.393*** | 1.443*** | 1.393*** | 1.381*** |
| *  Unemployed* | | | 1.571*** | 1.614*** | 1.544*** | 1.511*** |
| *Education* | | | | | | |
| *  Lower* | | | | Ref. | Ref. | Ref. |
| *  Intermediate* | | | | 1.049$^{NS}$ | 1.063$^{NS}$ | 1.053$^{NS}$ |
| *  Higher* | | | | 1.371*** | 1.398*** | 1.386*** |
| *Possession of an automobile* | | | | | 0.701*** | 0.736*** |
| *Single-mother household* | | | | | | 1.185*** |

*** significant at a 1% level; ** significant at a 5% level; * significant at a 10% level; $^{NS}$ not significant at a 10% level. Number of observations: 27,044.

Table 3: Logistic regression of unemployment probability (odds ratios)

To provide a more intuitive interpretation, marginal effects of neighbourhood type were computed for Models II, III and VI. These give the change in predicted probability of unemployment associated with a change of neighbourhood type on the average individual of our sample. Results indicate that living in a deprived neighbourhood in comparison with not deprived neighbourhoods, increases the unemployment probability of the average young adult by 7.3 percentage points in Model II, 6.3 points in Model III and 6.3 points in Model VI. Thus, adding an extensive set of parental characteristics to a model including only individual characteristics makes the estimated neighbourhood effect fall by 1 point. Moreover, estimating neighbourhood effects including only parental employment status (Model III) gives the same result as in the more comprehensive specification (Model VI). By way of comparison, the observed unemployment rate for young adults living with parents is 31% in deprived neighbourhoods and 19% in the rest of the agglomeration. The estimated marginal effect for Model VI indicates that 6.3 points of this gap (i.e. approximately 50%) are due to neighbourhood effects. The remaining would be due to the sorting of individuals with similar personal and parental characteristics into neighbourhoods.

## 5.5. Sensitivity to unobserved covariates

After having evaluated the influence of observed parental characteristics on estimated neighbourhood effects, it may be useful to test the sensitivity of the results to the endogeneity bias which results from the omission of an unobserved parental covariate which is correlated to both the probability of unemployment among young adults and parental residential choice. The approach used here is based on the method developed by Rosenbaum and Rubin (1983) and recently applied by Harding (2003) in the context of neighbourhood effects. The goal of this analysis is to assess how an unobserved binary covariate which affects both the probability of unemployment among young adults and the choice of parents to reside in a deprived or a non-deprived neighbourhood would alter the conclusions of this research about the magnitude and significance of neighbourhood effects. This is done by generating a series of unobserved binary variables $U$ that vary according to their degree of association with the neighbourhood dummy variable $DN$ and with the binary outcome measure $Y$. The degrees of association between $U$ and $Y$ and between $U$ and $DN$ are measured by parameters $k$ and $l$, both expressed in terms of odds ratios.

In practice, the method implemented consists of generating a binary variable $U$ sampled according to the following logistic model:

$$\text{Log}\left(\frac{\text{Prob}(U_i = 1)}{1 - \text{Prob}(U_i = 1)}\right) = \alpha + \kappa y_i + \lambda DN_i \qquad [12]$$

where $\text{Prob}(U_i=1)$ is the probability that the unobserved variable $U$ takes a value 1 for individual $i$, $y_i$ is a binary variable indicating whether $i$ is unemployed or not and $DN_i$ is a binary variable indicating whether $i$ resides in a deprived neighbourhood or not. In this formulation, $\kappa=\log(k)$ and $\lambda=\log(l)$, with $k$ and $l$ being *chosen* odds ratios that measure the strength of the association that is *imposed* by the researcher between $U$ and $Y$ and $U$ and $DN$ respectively. $\alpha$ is determined so that the overall prevalence of $U_i=1$ is 0.5. More precisely, the previous equation and the three imposed constraints (on the two sensitivity parameters $k$ and $l$ and the overall prevalence of $U_i=1$) are used to determine the proportion of $U_i=1$ in each one of the four subgroups defined by $DN$ and $Y$, i.e. $p_{11}=P(U_i=1|DN_i=1,y_i=1)$, $p_{10}=P(U_i=1|DN_i=1,y_i=0)$, $p_{01}=P(U_i=1|DN_i=0,y_i=1)$ and $p_{00}=P(U_i=1|DN_i=0,y_i=0)$. A random variable $U$ is then generated in each one of these subgroups, following a Bernoulli distribution of parameter $p_{11}$ for subgroup $(ND_i=1,y_i=1)$, $p_{10}$ for

subgroup ($ND_i=1,y_i=0$), etc. Once each individual has received a value for $U$, new estimates of neighbourhood effects are obtained by including $U$ in the unemployment probability model given by [11]. This is repeated for increasing values of the sensitivity parameters $k$ and $l$ in order to investigate what level of endogeneity bias (i.e. the strength of the association between $U$ and $Y$ and between $U$ and $DN$) would be needed to invalidate the results and render the estimated neighbourhood effects not significant.

Table 4 presents the estimated odds ratios associated with living in a deprived neighbourhood on unemployment probability, for two different models: (*i*) the model including only individual characteristics (and neighbourhood type) and no parental characteristics (as in Table 3's Model II) and (*ii*) the model including the full set of parental controls (as in Table 3's Model VI). In each panel of Table 4, the odds ratio in the extreme top-left cell (corresponding to $k$ and $l$ equals to one) gives the baseline odds ratio associated with living in a deprived neighbourhood, without introducing any amount of selection on unobservables. This odds ratio is 1.491 in Model II and 1.418 in Model VI. For each one of these models, an artificially created binary variable $U$ is included and a sensitivity matrix is obtained by varying the sensitivity parameters $k$ and $l$, which measure the associations of the unobserved parental characteristic $U$ with the neighbourhood type and the employment status.

| *Model II* | Sensitivity parameter $k$ | | | | | | |
|---|---|---|---|---|---|---|---|
| Sensitivity parameter $l$ | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 |
| 1.0 | $1.491^{***}$ | $1.485^{***}$ | $1.488^{***}$ | $1.482^{***}$ | $1.494^{***}$ | $1.503^{***}$ | $1.490^{***}$ |
| 1.5 | $1.496^{***}$ | $1.437^{***}$ | $1.391^{***}$ | $1.366^{***}$ | $1.352^{***}$ | $1.308^{***}$ | $1.319^{***}$ |
| 2.0 | $1.491^{***}$ | $1.397^{***}$ | $1.327^{***}$ | $1.301^{***}$ | $1.250^{***}$ | $1.217^{***}$ | $1.187^{***}$ |
| 2.5 | $1.494^{***}$ | $1.380^{***}$ | $1.274^{***}$ | $1.211^{***}$ | $1.200^{***}$ | $1.164^{***}$ | $1.123^{***}$ |
| 3.0 | $1.501^{***}$ | $1.365^{***}$ | $1.263^{***}$ | $1.184^{***}$ | $1.137^{***}$ | $1.112^{***}$ | $1.070^{**}$ |
| 3.5 | $1.506^{***}$ | $1.333^{***}$ | $1.229^{***}$ | $1.135^{***}$ | $1.093^{***}$ | $1.062^{*}$ | $1.032^{NS}$ |
| 4.0 | $1.514^{***}$ | $1.320^{***}$ | $1.205^{***}$ | $1.150^{***}$ | $1.073^{**}$ | $1.025^{NS}$ | $0.992^{NS}$ |

| *Model VI* | Sensitivity parameter $k$ | | | | | | |
|---|---|---|---|---|---|---|---|
| Sensitivity parameter $l$ | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 |
| 1.0 | $1.418^{***}$ | $1.411^{***}$ | $1.418^{***}$ | $1.412^{***}$ | $1.420^{***}$ | $1.433^{***}$ | $1.417^{***}$ |
| 1.5 | $1.423^{***}$ | $1.367^{***}$ | $1.325^{***}$ | $1.297^{***}$ | $1.287^{***}$ | $1.246^{***}$ | $1.256^{***}$ |
| 2.0 | $1.417^{***}$ | $1.328^{***}$ | $1.265^{***}$ | $1.237^{***}$ | $1.187^{***}$ | $1.163^{***}$ | $1.132^{***}$ |
| 2.5 | $1.420^{***}$ | $1.309^{***}$ | $1.212^{***}$ | $1.155^{***}$ | $1.143^{***}$ | $1.106^{***}$ | $1.069^{*}$ |
| 3.0 | $1.431^{***}$ | $1.300^{***}$ | $1.198^{***}$ | $1.126^{***}$ | $1.119^{***}$ | $1.085^{**}$ | $1.058^{NS}$ |
| 3.5 | $1.431^{***}$ | $1.265^{***}$ | $1.170^{***}$ | $1.077^{**}$ | $1.040^{NS}$ | $1.011^{NS}$ | $0.982^{NS}$ |
| 4.0 | $1.438^{***}$ | $1.256^{***}$ | $1.147^{***}$ | $1.093^{**}$ | $1.018^{NS}$ | $0.973^{NS}$ | $0.946^{NS}$ |

*** significant at a 1% level; ** significant at a 5% level; * significant at a 10% level; NS not significant at a 10% level. Number of observations: 27,044. Sensitivity parameters $k$ and $l$ are expressed in terms of odds ratios and measure the effect of an artificially created binary variable $U$ on the probability of unemployment and the probability of living in a deprived neighbourhood respectively.

Table 4: Sensitivity of neighbourhood effect estimates to the presence of an unobserved covariate (odds ratios)

As expected, Table 4 shows that in both specifications, the estimated neighbourhood effect decreases as the values of $k$ and $l$ increase. This means that accounting for a previously omitted variable correlated both

with the neighbourhood type and with the employment status does indeed reduce the intensity of neighbourhood effects estimates. However, the effect seems fairly robust since the values of $k$ and $l$ would have to be very high to make the neighbourhood effect not significant. Indeed, in the model including no parental or household characteristic (Model II), an unobserved covariate which multiplies the odds of living in a deprived neighbourhood by 3.5 and the odds of being unemployed by 4.0 would be required to totally erase the neighbourhood effect. This is also true when the full set of observed parental characteristics is included in the analysis (Model VI). In this case, the neighbourhood effect becomes insignificant when at least one of the two odds ratios reaches 3.5 and the other one equals 3.0.

In order to provide a more substantive interpretation, one may compare values for $k$ and $l$ (i.e. the amount of selection on unobservables that would be needed to make neighbourhood effects disappear) to the odds ratios estimated on observed covariates. Let assume for example that the unobserved variable $U$ is parental involvement with their children (this is indeed the most-often mentioned source of bias in neighbourhood effects studies; see O'Regan and Quigley, 1996; Harding, 2003). Though parental involvement is not measurable using standard census surveys, it is likely to affect both their children labour-market outcomes and their residential location. Indeed, whatever their income, social status or educational level, more involved parents will try to help their children as much as they can, by offering them support and advices in their job search effort or by monitoring their peer relations. At the same time, these involved parents may anticipate the detrimental effects of the social environment on their children and make some financial sacrifices to afford housing outside distressed areas. The question asked by this sensitivity analysis is then: "How big would the effect of parental involvement need to be to completely wipe out the effect of neighbourhood?"

The sensitivity analysis shows that having involved parents would need to increase the odds of being unemployed by 4.0 in order to totally erase the neighbourhood effect in Model II (or 3.0 in Model VI). As can be seen from Table 3, none of the observed parental and household characteristics already included in the model produces odds ratios as high as 4.0 or even 3.0, the highest odds ratio for a parental characteristics being 1.571 in Model VI (for having an unemployed parent). Among individual characteristics, the highest odds ratio is 2.374 for North-African or Turkish citizenship, and it is still below the 3.0 limit obtained by the sensitivity analysis. In other words, for the neighbourhood effect to become insignificant when considering parental involvement, the effect of this unobserved characteristic on unemployment would have to be stronger than that of parental employment status or that of citizenship, which seems not realistic. This provides some relatively strong evidence on the robustness of estimated neighbourhood effects.

The statistical analyses conducted here thus confirm that residential location does influence the labour-market outcomes of young adults residing with their parents in the Brussels agglomeration. Living in one of the deprived neighbourhoods of Brussels, i.e. characterized by the worst combination of social characteristics, significantly increases the unemployment probability of young adults. Although it may be feared that focusing on young adults residing with their parents creates some sample selection problems, the sensitivity analysis developed here suggests that estimated neighbourhood effects are robust in the presence of both observed and unobserved parental covariates. Indeed, the amount of selection on unobservables would have to be unreasonably high to make the estimated neighbourhood effect not significant.

## 6.    Conclusion

The objective of this paper was to investigate the endogeneity issue, which is one of the prominent problems encountered in neighbourhood effects studies. Indeed, despite the huge amount of empirical studies, there is still considerable debate about the existence and magnitude of neighbourhood effects (Ellen and Turner, 1997). This is mainly because empirical studies are subject to several methodological problems (Dietz, 2002, Durlauf, 2004), in particular to an endogeneity bias arising from the fact that individuals are not randomly distributed into the urban area but instead "self-selected" into neighbourhoods on the basis of their personal characteristics. In other words, some individual characteristics that influence individual outcomes (for example their employment status) also influence their residential choice. This means that what the researcher perceives as an effect of neighbourhood on individual outcomes may simply stem from a correlated effect reflecting common residential choice.

This paper contributes to the literature on neighbourhood effects and endogeneity issues by firstly reviewing the methods that have been proposed in the literature to try to solve the endogeneity issue. Secondly this paper shows (on the basis of one empirical example) how to cope with endogeneity and how to evaluate endogeneity biases in neighbourhood effects estimates. To this end, the effect of living in a deprived neighbourhood on the unemployment probability of young adults residing in Brussels is estimated by means of logistic regressions. The endogeneity of residential choices is addressed by means of the sample restriction method, which consists in restricting the sample to young adults residing with their parents. This is the simplest and most often used method in the literature (e.g. O'Regan and Quigley, 1996). It is based on the argument that residential choices have been made previously by the parents and can thus be considered as fairly exogenous to the employment status of their children. However, it is an imperfect solution as unobserved parental characteristics may still influence both the residential choice of parents and the employment status of their children (Ihlanfeldt and Sjoquist, 1998). In this context, the methodological originality of this paper is to evaluate the potential remaining endogeneity biases by conducting a sensitivity analysis to the presence of both observed and unobserved parental characteristics.

The results of this empirical application mainly showed that living in a deprived neighbourhood significantly increases the unemployment probability of young adults in Brussels, which confirm previous findings on 1991 data (Dujardin et al, 2008). This result is robust in the presence of both observed and unobserved parental covariates. Indeed, neighbourhood effects remain statistically significant when an extensive set of parental controls is introduced in the regression. Moreover, the sensitivity analysis based on artificially created unobserved covariates developed for this analysis shows that the amount of selection on unobservables would have to be unreasonably high to render neighbourhood effects estimates not significant.

However, the results also suggest that, while it seems that it is enough to remove much endogeneity biases, the sample restriction method presents other shortcomings. Indeed, besides the fact that it does not allow generalizing the results to other age groups, restricting the sample to young adults residing with parents is likely to generate sample selection problems (for example, if the leaving rate of young adults once they have found a job differ according to the characteristics of parents and/or according to the type of neighbourhood they live in). Some counter-intuitive findings of the statistical analyses conducted in this paper tend to indicate that this is probably the case in Brussels. In this context, it seems important that future research on neighbourhood effects focus on exploring other ways to solve the endogeneity issues.

As suggested by Durlauf (2004), the solution to endogeneity will probably have to be searched in the simultaneous modelling of neighbourhood effects and neighbourhood choices, through structural models. This will require a better understanding of residential choices as well as feedback mechanisms between individual outcomes and the characteristics of the social environment.

## Information

The sensitivity analysis to unobserved covariates developed in this paper makes use of the *Sensuc* function, which is part of the Design S library written by F. Harrell. Documentation and programs can be found at the following web links:
http://lib.stat.cmu.edu/S/Harrell/Design.html (general introduction to the Design library)
http://lib.stat.cmu.edu/S/Harrell/help/Design/html/sensuc.html (specific documentation on the Sensuc function)

## References

Aaronson, D., 1998, Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources*, 33, 4, pp. 915-946.

Benabou, R., 1993, Workings of a city: location, education and production. *Quarterly Journal of Economics*, 108, 3, pp. 619-652.

Blakely, T. Subramanian, S.V., 2006, Multilevel studies. In: Oakes, M. and Kaufman, J., editors, *Methods in Social Epidemiology*, pp. 316-340, Jossey Bass, San Francisco.

Blume, L.E., Durlauf, S.N., 2006, Identifying social interactions: a review. In: Oakes, M. and Kaufman, J., editors, *Methods in Social Epidemiology*, pp. 287-315. Jossey Bass, San Francisco.

Brock, W.A., Durlauf, S.N., 2001, Interaction-based models. In: Heckman, J.J. and Leamer, E.E., editors, *Handbook of Econometrics*, Volume 5, pp. 3297-3380. North-Holland: Amsterdam.

Brulard, T., Van der Haegen, H., 1972, La division des communes belges en secteurs statistiques. Le point de vue des géographes. *Acta Geographica Lovaniencia*, 10, pp. 21-36.

Buck, N., 2001, Identifying neighborhood effects on social exclusion. *Urban Studies*, 38, 12, pp. 2251-2275.

Case, A.C., Katz, L., 1991, The company you keep: the effects of family and neighborhood on disadvantaged youths. NBER Working Paper N°3705.

Corcoran, M., Gordon, R., Laren, D., Solon, G., 1992, The association between men's economic status and their family and community of origins. *The Journal of Human Resources*, 26, 4, pp. 575-601.

Crane, J., 1991, The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American Journal of Sociology*, 96, 5, 1226-1259.

Cutler, D.M., Glaeser, E.L.,1997, Are ghettos good or bad? *Quarterly Journal of Economics*, 112, pp. 827-872.

Datcher, L., 1982, Effects of community and family background on achievement. *The review of Economic and Statistics*, 64, 1, pp. 32-41.

Dietz, R.D., 2002, The estimation of neighborhood effects in the social sciences: an interdisciplinary approach. *Social Science Research*, 31, 4, pp. 539-575.

Dujardin, C., 2006, The role of residential location on labour-market outcomes. The urban agglomerations of Lyon and Brussels, PhD Thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

Dujardin, C., Selod, H., Thomas, I., 2004, Le chômage dans l'agglomération bruxelloise: une explication par la structure urbaine. *Revue d'Economie Régionale Urbaine*, 1, pp. 3-28.

Dujardin, C., Selod, H., Thomas, I., 2008, Residential segregation and unemployment: the case of Brussels. *Urban Studies*, 45, 1, pp. 89-113.

Duncan, G.J., Aber, J.L., 1997, Neighborhood models and measures. In: Brooks-Gunn, J. et al, editors, *Neighborhood Poverty. Volume 1: Context and Consequences for Children*, pp. 62-78. Russel Sage Foundation.

Duncan, G.J., Connell, J.P., Klebanov, P.K., 1997, Conceptual and methodological issues in estimating causal effects of neighborhoods and family conditions on individual development. In: Brooks-Gunn, J. et al, editors, *Neighborhood Poverty. Volume 1: Context and Consequences for Children*, pp. 219-249. Russel Sage Foundation.

Durlauf, S.N., 2004, Neighborhood effects. In: Henderson, J.V. and Thisse; J.-F., editors, *Handbook of Regional and Urban Economics*, Volume 4, pp. 2173-2242, North-Holland, Amsterdam.

Ellen, I.G., Turner, M.A., 1997, Does neighborhood matter? Assessing recent evidence. *Housing Policy Debate*, 8, 4, pp. 833-866.

Evans, W.N., Oates, W.E., Schwab, R.M., 1992, Measuring peer group effects: a study of teenage behavior, *Journal of Public Economics*, 100, 5, pp. 966-991.

Fieldhouse, E.A., Gould, M.I., 1998, Ethnic minority unemployment and local labour market conditions in Great Britain. *Environment and Planning A*, 30, 5, pp. 833-853.

Ginther, D., Haveman, R., Wolfe, B., 2000, Neighborhood attributes as determinants of children's outcomes. *The Journal of Human Resources*, 35, 4, pp. 603-642.

Glaeser, E.L., 1996, Spatial effects upon unemployment outcomes: the case of New Jersey teenagers. Discussion. *New England Economic Review*, May/June, pp. 58-64.

Gobillon, L., Selod, H., Zenou, Y., 2007, The mechanisms of spatial mismatch. *Urban Studies*, 44, 12, 2401-2427.

Goldstein, H., 2003, *Multilevel Statistical Models*. Edward Arnold, London.

Greene, W.H., 2008, *Econometric Analysis*. Sixth Edition. Prentice Hall.

Harding, D.J., 2003, Counterfactual models of neighborhood effects: the effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology*, 109, 3, pp. 676-719.

Holzer, H., 1988, Search method used by unemployed youth. *Journal of Labor Economics*, 6, 1, pp. 1-20.

Ihlanfeldt, K.R., Sjoquist, D.L., 1998, The spatial mismatch hypothesis: a review of recent studies and their implications for welfare reform. *Housing Policy Debate*, 9, 4, pp. 849-892.

Ioannides, Y.M., Zabel, J.E., 2008, Interactions, neighborhood selection and housing demand. *Journal of Urban Economics*, 63, 229-252.

Jencks, C., Mayer, S.E., 1990, The social consequences of growing up in a poor neighbourhood. In: Lynn, L.E. and McGeary, M.G.H., editors, *Inner-city Poverty in the United States*, pp. 111-186. National Academy Press, Washington D.C.

Johnston, R., Jones, K., Burgess, S., Propper, R., Sarker, R., Bolster, A., 2004, Scale, factor analyses, and neighborhood effects. *Geographical Analysis*, 36, 4, pp. 350-368.

Katz, L.F., Kling, J.R., Liebman, J.B., 2001, Moving to Opportunity in Boston: early results of a randomized mobility experiment. *Quarterly Journal of Economics*, 116, 2, pp. 607-654.

Kling, J.R., Liebman, J.B., Katz, L.F., 2007, Experimental analysis of neighborhood effets. *Econometrica*, 75, 1, 83-119.

Manski, C.F., 1993, Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, 60, pp. 531-543.

Manski, C.F., 1995, *Identification Problems in the Social Sciences*. Harvard University Press, Chicago.

Oakes, J.M., 2004, The (mis)estimation of neighbourhood effects: causal inference for a practical social epidemiology. *Social Science and Medicine*, 58, 10, pp. 1929-1952.

O'Regan, K.M., Quigley, J.M., 1996, effects upon unemployment outcomes: the case of New Jersey teenagers. *New England Economic Review*, May/June, pp. 41-58.

Oreopoulos, P.,2003, The long-run consequences of living in a poor neighborhood. *Quarterly Journal of Economics*, 118, 4, pp. 1533-1575.

Osterman, P., 1991, Welfare participation in a full employment economy: the impact of neighborhood. *Social Problems*, 32, 4, pp. 475-491.

Plotnick, R.D., Hoffman, S.D., 1999, The effect of neighborhood characteristics on young adult outcomes: alternative estimates. *Social Science Quarterly*, 80, 1, pp. 1-18.

Reingold, D., 1999, Social networks and the employment problem of the urban poor. *Urban Studies*, 35, 11, pp. 1907-1932.

Robinson, W.S., 1950, Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, pp. 351-357.

Rosenbaum, P.R., Rubin, D.B., 1983, Assessing the sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society B*, 45, 2, pp. 212-218.

Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T., 2002, Assessing neighborhood effects: social processes and new directions in research. *Annual Review of Sociology*, 28, pp. 443-478.

Weinberg, B.A., Reagan, P.B., Yankow, J.J., 2004, Do neighborhoods affect work behaviour? Evidence from the NLSY79. *Journal of Labor Economics*, 22, 4, pp. 891-924.

Wilson, W.J., 1987, *The Truly Disadvantaged: the Inner City, the Underclass, and Public Policy.* The University of Chicago Press, Chicago.

Wong, D.W.S., 2004, Comparing traditional and spatial segregation measures: a spatial scale perspective. *Urban Geography*, 25, 1, pp. 66-82.

Wrigley, N., 1995, Revisiting the MAUP and the ecological fallacy. In: Cliff, A.D. et al, editors, *Diffusing Geography Essays for Peter Hagget.* The Institute of British Geographers, Special Publications Series.

Zenou, Y., Boccard, N., 2000, Labor discrimination and redlining in cities. *Journal of Urban Economics*, 48, 2, pp. 260-285.

# Recent titles

## CORE Discussion Papers

2009/18.    David DE LA CROIX and Michel LUBRANO. The tradeoff between growth and redistribution: ELIE in an overlapping generations model.
2009/19.    Thierry BRECHET and Fabien PRIEUR. Can education be good for both growth and the environment?
2009/20.    Giacomo SBRANA and Andrea SILVESTRINI. What do we know about comparing aggregate and disaggregate forecasts?
2009/21.    Marc GERMAIN, Henry TULKENS and Alphonse MAGNUS. Dynamic core-theoretic cooperation in a two-dimensional international environmental model.
2009/22.    Claude D'ASPREMONT and Rodolphe DOS SANTOS FERREIRA. Household behavior and individual autonomy.
2009/23.    Helmuth CREMER, Philippe DE DONDER and Pierre PESTIEAU. Education and social mobility.
2009/24.    Maria Eugenia SANIN and Francesco VIOLANTE. Understanding volatility dynamics in the EU-ETS market: lessons *from* the future.
2009/25.    Marco DI SUMMA and Laurence A. WOLSEY. Lot-sizing with stock upper bounds and fixed charges.
2009/26.    Johanna M.M. GOERTZ and François MANIQUET. On the informational efficiency of simple scoring rules.
2009/27.    Jean GABSZEWICZ, Ornella TAROLA and Skerdilajda ZANAJ. On uncertainty when it affects successive markets.
2009/28.    Jerzy A. FILAR, Jacek B. KRAWCZYK and Manju AGRAWAL. On production and abatement time scales in sustainable development. Can we loosen the *sustainability screw* ?
2009/29.    María Eugenia SANIN and Skerdilajda ZANAJ. Clean technology adoption and its influence on tradeable emission permit prices.
2009/30.    Antoine BOMMIER, Marie-Louise LEROUX and Jean-Marie LOZACHMEUR. On the public economics of annuities with differential mortality.
2009/31.    Gilles GRANDJEAN, Ana MAULEON and Vincent VANNETELBOSCH. Connections among farsighted agents.
2009/32.    Axel GAUTIER and Xavier WAUTHY. On the nature of price competition under universal service obligations: a note.
2009/33.    Santanu S. DEY and Laurence A. WOLSEY. Constrained infinite group relaxations of MIPs.
2009/34.    Jean-François MAYSTADT and Philip VERWIMP. Winners and losers among a refugee-hosting population.
2009/35.    Pierre DEHEZ. Allocation of fixed costs and the weighted Shapley value.
2009/36.    Sabien DOBBELAERE, Roland Iwan LUTTENS and Bettina PETERS. Starting an R&D project under uncertainty.
2009/37.    Carlotta BALESTRA and Davide DOTTORI. Aging society, health and the environment.
2009/38.    Alain PHOLO BALA. Urban concentration and economic growth: checking for specific regional effects.
2009/39.    Alain PHOLO BALA. Gates, hubs and urban primacy in Sub-Saharan Africa.
2009/40.    Nicolas BOCCARD. On efficiency, concentration and welfare.
2009/41.    Taoufik BOUEZMARNI, Jeroen V.K. ROMBOUTS and Abderrahim TAAMOUTI. A nonparametric copula based test for conditional independence with applications to Granger causality.
2009/42.    Josez KONINGS and Hylke VANDENBUSSCHE. Antidumping protection hurts exporters: firm-level evidence from France.
2009/43.    Pierre PESTIEAU and Uri M. POSSEN. Retirement as a hedge.
2009/44.    Santanu S. DEY and Laurence A. WOLSEY. Lifting group inequalities and an application to mixing inequalities.
2009/45.    Jean CAVAILHES, Pierre FRANKHAUSER, Dominique PEETERS and Isabelle THOMAS. Residential equilibrium in a multifractal metropolitan area.

# Recent titles

## CORE Discussion Papers - continued

2009/46.    Daisuke OYAMA, Yasuhiro SATO, Takatoshi TABUCHI and Jacques-François THISSE. On the impact of trade on industrial structures: The role of entry cost heterogeneity.
2009/47.    Ken-Ichi SHIMOMURA and Jacques-François THISSE. Competition among the big and the small.
2009/48.    Frédéric BABONNEAU, Yurii NESTEROV AND Jean-Philippe VIAL. Design and operations of gas transmission networks.
2009/49.    Olivier BOS. How lotteries outperform auctions for charity.
2009/50.    Nicolas BOCCARD and Xavier WAUTHY. Entry accommodation under multiple commitment strategies: judo economics revisited.
2009/51.    Joachim GAHUNGU and Yves SMEERS. Multi-assets real options.
2009/52.    Nicolas BOCCARD and Xavier WAUTHY. Regulating quality by regulating quantity: a case against minimum quality standards.
2009/53.    David DE LA CROIX and Frédéric DOCQUIER. An incentive mechanism to break the low-skill immigration deadlock.
2009/54.    Henry TULKENS and Vincent VAN STEENBERGHE. "Mitigation, adaptation, suffering": In search of the right mix in the face of climate change.
2009/55.    Santanu S. DEY and Quentin LOUVEAUX. Split rank of triangle and quadrilateral inequalities.
2009/56.    Claire DUJARDIN, Dominique PEETERS and Isabelle THOMAS. Neighbourhood effects and endogeneity issues.

## Books

*Public goods, environmental externalities and fiscal competition: 22 selected papers in public economics by Henry Tulkens*, edited and introduced by Parkash Chander, Jacques Drèze, C. Knox Lovell and Jack Mintz, Springer, Boston 2006 (588 pp.).

V. GINSBURGH and D. THROSBY (eds.) (2006), *Handbook of the economics of art and culture.* Amsterdam, Elsevier.

J. GABSZEWICZ (ed.) (2006), La différenciation des produits. Paris, La découverte.

L. BAUWENS, W. POHLMEIER and D. VEREDAS (eds.) (2008), *High frequency financial econometrics: recent developments*. Heidelberg, Physica-Verlag.

P. VAN HENTENRYCKE and L. WOLSEY (eds.) (2007), *Integration of AI and OR techniques in constraint programming for combinatorial optimization problems*. Berlin, Springer.

P-P. COMBES, Th. MAYER and J-F. THISSE (eds.) (2008), *Economic geography: the integration of regions and nations*. Princeton, Princeton University Press.

J. HINDRIKS (ed.) (2008), *Au-delà de Copernic: de la confusion au consensus* ? Brussels, Academic and Scientific Publishers.

J-M. HURIOT and J-F. THISSE (eds) (2009), *Economics of cities.* Cambridge, Cambridge University Press.

## CORE Lecture Series

C. GOURIÉROUX and A. MONFORT (1995), Simulation Based Econometric Methods.

A. RUBINSTEIN (1996), Lectures on Modeling Bounded Rationality.

J. RENEGAR (1999), A Mathematical View of Interior-Point Methods in Convex Optimization.

B.D. BERNHEIM and M.D. WHINSTON (1999), Anticompetitive Exclusion and Foreclosure Through Vertical Agreements.

D. BIENSTOCK (2001), Potential function methods for approximately solving linear programming problems: theory and practice.

R. AMIR (2002), Supermodularity and complementarity in economics.

R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.