MYRA YAZBECK

THREE ESSAYS IN HEALTH ECONOMICS

Thèse présentée
à la Faculté des études supérieures et postdoctorales de l'Université Laval
dans le cadre du programme de doctorat en économique
pour l'obtention du grade de Philosophiae Doctor (Ph.D.)

DÉPARTEMENT D'ÉCONOMIQUE
FACULTÉ DES SCIENCES SOCIALES
UNIVERSITÉ LAVAL
QUÉBEC

2011

Résumé

L'objectif principal de cette thèse est de comprendre comment les interactions sociales peuvent affecter l'état de santé via des changements dans certains comportements. Trois raisons motivent l'intérêt pour cette question. Premièrement, des habitudes de vie telles que les mauvaises habitudes alimentaires ou le manque d'activité physique représentent un facteur de risque important pour la santé et donc une source de coûts qui pourraient être évités. Deuxièmement, il s'agit de mieux comprendre la façon dont certains problèmes affectent le capital santé des jeunes. Troisièmement, les interactions sociales qui influencent certains comportements sont une source d'externalités qui amplifient l'impact de chocs sur la santé, nécessitant ainsi l'adoption de politiques appropriées.

Dans le premier essai, nous analysons l'épidémie d'obésité observée chez les jeunes américains. Ce problème a fait couler beaucoup d'encre récemment. La plupart des recherches publiées sur ce sujet étudient l'aspect socialement contagieux de l'obésité mais sans en expliquer les mécanismes. Dans un premier temps, nous allons au delà de cette approche pour analyser si la présence d'une telle épidémie sociale est due à des effets de pairs dans les habitudes alimentaires. Nous mettons en particulier l'accent sur la consommation de malbouffe. Dans un deuxième temps, nous examinons l'impact de cette consommation sur la dynamique des gains de poids.

Dans le deuxième essai, nous nous intéressons à un second mécanisme de contagion sociale, soit les interactions sociales dans la pratique d'activités physiques chez les jeunes américains. A cet effet, nous évaluons d'abord la présence d'effets de pairs sur ces comportements chez les adolescents. Nous estimons ensuite l'impact de l'activité physique sur leur état de santé.

Le troisième essai porte sur l'impact d'un régime relativement généreux de pension de vieillesse sur la santé des enfants en Afrique du Sud. Nous étudions en particulier si les effets attribués à ce programme sont dus à des changements comportementaux se traduisant par une augmentation de la cohabitation entre personnes Egées prestataires de la pension et les jeunes enfants.

Abstract

This thesis is composed of three essays. The first paper aims at opening the black box of peer effects in adolescent weight gain. Using Add Health data on secondary schools in the U.S., I investigate whether these effects partly flow through the *eating habits* channel. Adolescents are assumed to interact through a friendship social network. I first propose a social interaction model of fast food consumption. My approach allows to control for correlated effects at the network level and to solve the simultaneity (reflection) problem. I exploit results by Bramoullé, Djebbari and Fortin (2009) which show that if there are two agents who are separated by a link of distance 3 within a network (*i.e.*, two adolescents who are not friends but are linked by two friends), peer effects are identified. The model is estimated using maximum likelihood and generalized 2SLS strategies. I also estimate a panel dynamic weight gain production function relating an adolescent's Body Mass Index (BMI) to his current fast food consumption and his lagged BMI level. Results show that there are positive significant peer effects in fast food consumption among adolescents belonging to a same friendship school network. The estimated social multiplier is 1.59. My results also suggest that, at the network level, an extra day of weekly fast food restaurant visits increases BMI by 2.4%, when peer effects are taken into account.

The objective of the second paper is to assess wether there are peer effects in physical activity in an attempt to shed light on mechanisms through which peer effects in health outcomes may flow. I first propose a social interaction model for adolescents' physical activity. In this model adolescents interact through a friendship social network. The approach that I use accounts for correlated effects at the network level and solves the simultaneity problem. Using the identification method developed by Bramoullé et al. (2009) I estimate the peer effect model using maximum likelihood and generalized 2SLS. I also estimate a dynamic health production function in which I related an individual health status to his physical activity and his lagged health status. Estimation results show that there are positive significant peer effects in exercising of 0.11 and that exercising has a positive significant effect on health status of 0.09. The associated social multiplier is 1.12. This suggests that an increase in physical activity frequency improves health status by 5 % in the short-run and 6% in the long-run.

Finally, based on the empirical evidence on the impact of South African old-age pension on children's health (Duflo, 2003) as well as on living arrangements (Edmonds et al., 2005) the third paper investigates wether this pension program may be affecting

children's health through changes in living arrangements. Most of the literature on the income-health gradient focuses on the challenges to causal inference arising from omitted variable bias and simultaneity problem, yet few tackle the mechanisms by which income affects health. This paper contributes to the literature on the mechanisms of the income-health gradient by extending the work of Case and Deaton (1998) and Duflo (2003) and exploiting results from the statistical and econometric literature (Flores and Flores-Lagunes, 2007) to identify the effect of the pension income on children's health net of changes in living arrangements. My results show that there is no evidence of a net income effect. This suggests the results obtained previously may be entirely due to changes in the compositions of households with no real improvement in children's health. It also suggests that pension cash transfers targeted to the elderly poor do not trickle down to children. It cautions against conditioning on variables that are themselves affected by a treatment when estimating average treatment effect on beneficiaries.

Acknowledgements

I would like to thank:

My thesis director M. Bernard Fortin for his precious help, guidance and availability during the last six years. I would also like to thank my thesis co-director Mrs. Habiba Djebbari for her time and advice. It is real pleasure to interact and work with both of you. A special thanks to Charles Bellemare and Guy Lacroix for their help and support, I really appreciate it.

The external examiner Pierre-Carl Michaud for his useful comments.

The Canada Research Chair in the Economics of Social Policies and Human Resources and le Centre interuniversitaire sur le risque, les politiques économiques et l'emploi for the financial support.

My daughter Karen who bared with me all through my graduate studies.

My parents Akram and Georgette as well as my brother Adib for their overseas encouragement and support.

My family in Québec Paul, Etienne, Jeanne, Marie-Thérèse and Marie-Louise the journey was much easier in your presence and with your support.

Lynda Khalaf who has been my mentor, my professor and my friend and who managed to keep each of these roles independent.

My friends Daniella, Hiam and Elham for being actively present in my life and for taking care of Karen when I needed to work.

Last but not least, I would like to thank all my professors, all the members of the economics department at Université Laval and all my friends in Québec city. I cherish every moment I spent with each and every one of you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation focuses on modelling the effects of social networks (friends and household structure) on health. Externalities created by social networks have recently gained interest and credibility in public and health economics. In this context, leading research questions from a public economic perspective concern the pathways through which interactions of networks and health can be channeled. My thesis provides an empirical analysis of such interaction channels, in developed and developing countries. In three essays, I analyze various health outcomes including adolescents' eating habits, physical activity as well as children anthropometric measures, and identify different interaction and transmission mechanisms. From a public economic perspective, my focus on pre-adults (adolescents and children) is motivated by policy considerations. Childhood and adolescence investment in health capital (e.g., lifestyle, nutrition) sets the ground for long term health. Indeed, public policies that affect youths' health capital may be more efficient than alternative redistributive welfare policies targeting adult stages of life.

From this perspective, I first investigate the prevalent problem of obesity among young adolescents in the United States. Lately, the economic burden of obesity costs has been one of the major concerns of health policy makers in the United States. While there are many theories by which one may explain this spread, my approach is based on the hypothesis that an exogenous technological shock on fast food supply accentuated by the presence of peer effects at the consumption level may be behind this spread. The increase in fast food supply lead to an increase the quantity of fast food consumed. In the presence of a social multiplier, this increase lead to an increase in the demand for fast food and thus to a further increase in the quantities consumed.

The empirical evidence provided in the economic literature seems to be pointing at the social multiplier as a responsible for the obesity epidemics (Christakis and Fowler,

2007a; Trogdon et al., 2008; Renna et al., 2008). Yet, the validity evidence provided by Christakis and Fowler (2007a) has been debated by Cohen-Cole and Fletcher (2008a) and Fowler and Christakis (2008c). Recently this debate was revived and the approach adopted by Christakis and Fowler (2007a) was criticized Lyons (2011) and Shalizi and Thomas (2011). While the statistical methodologies used in the estimation of social interaction models are still not perfect, I use one of the most recent identification and estimation methods.

In my second chapter, I take the evidence on the presence of peer effect and obesity a step ahead, and go beyond the black box approach to peer effects models. I identify one of the possible behavioural channels through which the identified peer effect may flow: eating habits. A simple interaction with an obese person is unlikely to produce a direct impact on a person's body weight. Thus, if such contagion exists, then it must be mediated at the behavioural level. This behavioural change may occur as a result of a direct effect (i.e., emulation) on eating habits. It may also be the result of an indirect effect namely through a change weight perceptions (Blanchflower and Oswald, 2009; Oswald and Powdthavee, 2007). In this paper, I focus on the behavioural mechanism as it presents many advantages. First, even if there are peer effects at the perception level these peer effects are translated at the behavioural level. By investigating the presence of peer effects in eating habits (more specifically fast food consumption) one can capture this indirect dimension. Second, by looking at a behavioural variable (i.e., eating habits) rather than an outcome variable (obesity), one can use the derived information for policy purposes. Policies targeting eating habits are easier to implement than policies targeting individual's weight (which are discriminatory). Finally, since it is much easier to eat than to exercise, one can suspect that eating habits may be the leading cause of weight gain.

To highlight the presence of a behavioural mechanism beneath the obesity epidemic, I use a two equation model. In the first equation, I use the linear-in-means model to model the peer effects in eating habits (proxied by fast food consumption).[1] There are well known identification issues attributed to the use of such a model (Manski, 1993). To address these identification issues, I use the recently developed methods by Bramoullé et al. (2009). According to the authors, once the correlated effects are taken into account, it is possible to identify endogenous peer effects as long as one can find two strangers who are related by two friends in the same social network. Using data from Add-Health, I consider peers (nominated friends) within the same school and test if the identification condition is satisfied. Then, I exploit the similarities between the linear in means model and spatial autoregressive model to estimate endogenous peer

[1]De Agostini et al. (2007) provides evidence that suggests that rising weight trends are due to calories consumed outside home rather than at home.

effects parameters utilizing maximum likelihood (as in Lee et al., 2010; Lin, 2010) and generalized spatial two stage least squares. As for the second equation in the model, it allows me to relate fast food consumption to weight gain through a production function.

Combining these two equations together, I can obtain the impact of a fast food consumption's exogenous shock on weight gain while accounting for the impact of the social multiplier. Basically, this chapter's results are consistent with the hypothesis that if there are peer effects in obesity, a non-negligible part of it flows through eating habits. This, in turn, suggests that policies that would tax "fast food" may result in a double dividend situation where the externalities produced by fast food consumption would be internalized and government revenues would increase.

In addition to the presence of peer effects in obesity, another body of the peer effect literature has investigated the presence peer effects in different health outcomes (Cohen-Cole and Fletcher, 2008a). This line of research was first used to show that one can find peer effects in any health outcome. However, a recently published article by Carrell et al. (2011) documented the presence of peer effects in physical fitness. The transmission mechanism behind these peer effects were not investigated by the authors. Yet, works by Yakusheva et al. (2010), Jago et al. (2010) and Babcock and Hartman (2010) using experimental approach, provided evidence that supports the presence of peer effects in exercising. In my third chapter, I combine experimental evidence on the presence of peer effects in health outcomes with the experimental evidence on the presence of peer effects in physical activity. I investigate whether the peer effects in active lifestyle can be a possible channel by which one can explain the presence of such peer effects in health outcomes. Why is it important to identify the mechanism? While assessing the presence of peer effects in physical activity is important *per se*, the associated social multiplier can become more useful (from a policy perspective) once related to an outcome of public interest.

To assess wether the presence of peer effects in health outcomes flow through peer effects in physical activity, I use Add-Health data on three physical activities: Exercising, active sports and recreational sports and follow closely the methodology that was used in chapter 2. This chapter's results provide evidence that allow for the conclusion that active lifestyle (namely exercising) is a mechanism through which peer effects in health status flow. Thus subsidizing physical activity may be a good avenue to promote healthier and more productive individuals as well as curb the spread of chronic disease in general.

My interest in the interaction of social networks (friends, family...) with health and the externalities it produces is not limited to developed countries. I investigate this

same interaction from another perspective: whether household composition may be a possible channel through which income may flow to affect children's health. To assess this possibility, I exploit the unique nature of a South African quasi-natural experiment (universal Old Age Pension) and use the longitudinal nature of the data to identify changes in households' composition. I then use this information to block the impact of the pension on household composition and identify the net impact of the program on children anthropometric measures. My findings suggest that the pension program affects health through the change in household composition which results in sorting of children according to their health status. This suggests that pension cash transfers targeted to the elderly poor do not always trickle down to children. It cautions against conditioning on variables that are themselves affected by a treatment when estimating an average treatment effect on beneficiaries.

This thesis consists of five chapters and unfolds as follows. Chapter two investigates whether the presence of peer effects in obesity partially flows through the presence of peer effects in fast food consumption. Chapter three looks at the presence of peer effects in physical activity in an attempt to highlight the presence of a transmission mechanism through which peer effects in health status flows. Chapter four, looks at the net impact of the South African old age pension on children health after netting out changes in their family composition. Chapter five concludes.

Chapter 2

Peer Effects, Fast Food Consumption and Adolescent Weight Gain

[1]

2.1 Introduction

For the past few years, obesity has been one of the major concerns of health policy makers in the U.S. It has also been one of the principal sources of increased health care costs. In fact, the increasing trend in children's and adolescents' obesity has raised the annual obesity-related hospital costs for this part of the population to $127 million per year. Obesity is also associated with increased risk of reduced life expectancy as well as with serious health problems such as type 2 diabetes (Ford et al., 1997; Maggio and Pi-Sunyer, 2003), heart disease (Li et al., 2006a; Calabr et al., 2009) and certain cancers (Abu-Abid et al., 2002; Calle, 2007), making obesity a real public health challenge.

Recently, a growing body of the health economics literature has tried to look into the obesity problem from a new perspective using a social interaction framework. The evidence suggests the presence of strong peer effects in weight gain. Christakis and Fowler (2007b), Trogdon et al. (2008) and Renna et al. (2008) seem to be unanimously

[1]This chapter was co-written with Bernard Fortin.

pointing to the *social multiplier* as an important element in the obesity epidemics.[2] A social multiplier may amplify, at the aggregate level, the impact of any shock that affects obesity at the individual level. This is so because the aggregate effect incorporates, in addition to the sum of the individual direct effects, positive indirect peer effects stemming from social interactions.[3]

While the presence of the social multiplier in weight gain has been widely researched, the literature on the mechanisms by which this multiplier flows is still scarce. Indeed, most of the relevant literature attempts to estimate the relationship between variables such as an individual's Body Mass Index (BMI) and his average peers' BMI, without exploring the channels at source of this potential linkage.[4] The aim of this paper is to go beyond the black box approach of peer effects in weight gain and try to identify one crucial mechanism through which peer effects in adolescence overweight may flow: *eating habits* (as proxied by fast food consumption).[5]

Three reasons justify our interest in eating habits in analyzing the impact of peer effects on teenage weight gain. First of all, there is important literature that points to eating habits as an important component in weight gain (*e.g.*, Levitsky et al., 2004; Niemeier et al., 2006; Rosenheck, 2008).[6] Second, one suspects that peer effects in eating habits are likely to be important in adolescence. Indeed, at this age, youngsters have increased independence in general and more freedom as far as their food choices are concerned (Rolfes and Whitney, 1996). Usually vulnerable, they often compare themselves to their friends and may alter their choices to conform to the behaviour of their peers. Therefore, unless we scientifically prove that obesity is a virus,[7] it is counter intuitive to think that one can directly gain weight by simply interacting with an obese person.[8] This is why we are inclined to think that the presence of real peer effects in weight gain can be estimated using behavioural channels such as eating habits.

[2]While Cohen-Cole and Fletcher (2008b) found that there is no evidence of peer effects in weight gain, most of the literature is consistent with the presence of these effects.

[3]Researchers also found that friends had a significant effect in adolescent smoking, drinking, illicit drug use (Clark and Loheac, 2007; Fletcher and Ross, 2011) and risky sex (Jaccard et al., 2005).

[4]One recent exception is Yakusheva et al. (2010) who look at peer effects in weight gain and in weight management behaviours such as eating and physical exercise, using randomly assigned pairs of roommates in freshman year.

[5]Another potentially important channel is physical activity (Trogdon et al., 2008).

[6]An indirect evidence of the relationship between eating habits and weight gain come from the literature on the (negative) effect of fast food prices on adolescent BMI (see Chou et al., 2005; Powell et al., 2007; Auld and Powell, 2008; Powell and Bao, 2009). See also Cutler et al. (2003) which relates the declining relative price of fast food and the increase in fast food restaurant availability over time to increasing obesity in the U.S.

[7]We acknowledge that some recent studies have pointed that obesity might be partially due to a virus ad-36 (see Rogers et al., 2007).

[8]Of course, having obese peers may influence a student's tolerance for being obese and therefore his weight management behaviours (Blanchflower and Oswald, 2009; Oswald and Powdthavee, 2007).

Third, our interest in peer effects in youths' eating habits is policy driven. There has been much discussion on implementing tax policies to address the problem of obesity (*e.g.*, Jacobson and Brownell, 2000; Marshall, 2000; Kim and Kawachi, 2006; Caraher and Cowburn, 2007). As long as peer effects in fast food consumption is a source of externality that may stimulate overweight among adolescents, it may be justified to introduce a consumption tax on fast food. The optimal level of this tax will depend, among other things, on the social multiplier of eating habits, and on the causal effect of fast food consumption on adolescent weight.

In order to analyze the impact of peer effects in eating habits on weight gain, we propose a two-equation model. The first equation relates the teenager's fast food consumption to his reference group's mean fast food consumption (*endogenous peer effect*), his individual characteristics, and his reference group's mean characteristics (*contextual peer effects*). This *linear-in-means* equation also provides an estimate of the social multiplier in fast-food consumption. The second equation is a panel dynamic production function that relates the teenager's BMI to his current fast food consumption and his lagged BMI level. The system of equations thus allows us to evaluate the impact of an eating habits' exogenous shock on weight gain, when peer effects on fast food consumption are taken into account.

Estimating our system of equations raises serious econometric problems. It is well known that the identification of peers effects (first equation) is a challenging task. These identification issues were pointed out by Manski (1993) and discussed among others by Moffit (2001), Brock and Durlauf (2001), and Blume et al. (2010). On one hand, (endogenous + contextual) peer effects must be identified from *correlated* (or confounding) factors. For instance, students in a same friendship group may have similar eating habits because they are alike or face a common environment. On the other hand, simultaneity between an adolescent's and his peers' behaviour (referred to as *the reflection problem* by Manski) makes it difficult to identify separately the endogenous peer effect and the contextual effects. This later task is important since the endogenous peer effect is the only source of a social multiplier.

We use a new approach based on Bramoullé, Djebbari and Fortin (2009) and Lee, Liu and Lee (2010) to address these identification problems and to estimate the peer effects equation. First, we assume that in their fast food consumption decisions, adolescents interact through a *friendship network*. Each school is assumed to form a network. School fixed effects are introduced to capture correlated factors associated with network invariant unobserved variables (e.g., similar preferences due to self-selection in schools, same school nutrition policies, distance from fastfood restaurants). The structure of friendship links within a network is allowed to be stochastic and endogenous but is

strictly exogenous, conditional on the school fixed effects and observable individual and contextual variables. To solve the reflection problem, we exploit results by Bramoullé et al. (2009) who show that if there are at least two agents who are separated by a link of distance 3 within a network (*i.e.*, there are two adolescents in a school who are not friends but are linked by two friends), both endogenous and contextual peer effects are identified. The intuition is quite simple when there are no school fixed effects. In this case, identification conditions are less restrictive: peer effects are identified when a friend of an adolescent's friend is not his friend (link of distance 2, or intransitive triad). This provides exclusion restrictions in the model. More specifically, the friends' friends mean characteristics can be used as instruments for the friends' mean fast food consumption. Finally, we exploit the similarity between the linear-in-means model and the spatial autoregressive (SAR) model. The model is estimated using a maximum likelihood (ML) approach as in Lee et al. (2010) and Lin (2010). We also estimate the model with a distribution free approach: generalized spatial two-stage least square (GS-2SLS) proposed in Kelejian and Prucha (1998) and refined in Lee (2003).

The estimation of the dynamic weight gain production function (second equation) also raises some challenging issues as fast food consumption and past BMI level are likely to be endogenous variables. In order to deal with these problems, we follow Arellano and Bond (1991) by first differencing the equation to eliminate the individual effects and by exploiting the orthogonality conditions that exist between lagged values of BMI and the current disturbances to generate instruments.

To estimate our model, we use three waves of the National Longitudinal Study of Adolescent Health (Add Health). We define peers as the nominated group of individuals reported as friends within the same school. The consumption behaviour is depicted through the reported frequency (in days) of fast food restaurant visits in the past week. Results suggest that there is a positive significant peer effect in fast food consumption among adolescents in general. Based on the ML approach, the estimated social multiplier is 1.59. Moreover, the production function estimates indicate that there is a positive significant impact of fast food consumption on BMI. Combining these results, we find that, at the network level, an extra day of fast food restaurant visits per week increases BMI by 2.4% on average within a year.

The remaining parts of this paper will be laid out as follows. Section 2 provides a critical survey of the literature on peer effects in obesity as well as its decomposition into the impact of peer effects on fast food consumption and the impact of fast food consumption on obesity. Section 3 presents our two-equation model and our estimation methods. In section 4, we give a brief overview of the Add Health Survey and we provide descriptive statistics of the data we use. In section 5, we discuss estimation

results. Section 6 concludes.

2.2 Previous literature

In recent years, a number of studies found strong "social networks effects" in weight outcomes. In a widely debated article, Christakis and Fowler (2007b), using a 32-year panel dataset on adults from Framingham, Massachusetts and based on a logit specification, found that an individual's probability of becoming obese increased by 57% if he or she had a friend who became obese in a given interval. However, their analysis has been criticized for suffering from a number of limitations (see Cohen-Cole and Fletcher, 2008b; Lyons, 2011; Shalizi and Thomas, 2011).[9] In particular, it ignores potential spurious correlations between two friends' BMI resulting from the fact that they are exposed to a same environment. Both Shalizi and Thomas (2011) and Lyons (2011) show that the relying on link asymmetries does not rule out shared environment as it claims. Also, the simultaneity problem between these two outcomes is not directly addressed by allowing the peer's obesity to be endogenous. Moreover, by introducing lagged obesity variables, it only partly takes into account the problem of selection that may occur as obese individuals may have a higher probability to become friends (for details see pages 217-218 in Shalizi and Thomas, 2011). Finally, by focusing on dyads over time, it introduces an upward bias resulting from the *unfriending* problem as defined by Noel and Nyhan (2011).[10] The basic idea behind this argument is that people who are alike (homophilious) are more likely to maintain social ties.[11]

Using Add health data, Trogdon et al. (2008) include school fixed effects to account for the fact that students in a same school share a same surrounding. They also estimate their BMI peer model with an instrumental variable approach. They use information on friends' parents' obesity and health and friends' birth weight as instruments for peers' BMI. They find that a one point increase in peers' average BMI increases own BMI by 0.52 point. Using a similar approach and based on Add Health dataset, Renna et al. (2008) also find positive peer effects. These effects are significant for females only (= 0.25 point). These analyses raise a number of concerns though. In particular, they assume no contextual variables reflecting peers' mean characteristics. This rules out the reflection problem by introducing non-tested restriction exclusions. As a result, the

[9]For a response to these criticisms and others, see Fowler and Christakis (2008b), Christakis and Fowler (2010) and VanderWeele (2011).

[10]Unfriending means people who stop being friends with each other.

[11]Steeg and Galstyan (2011) show that there is a test for ruling out homophily related to the use of longitudinal social networks. The intuition behind the test being that if an individual reproduces the same sequence of events as his friend, it is unlikely that homophily is a source of this replication.

peer effect estimates may be inconsistent. Moreover, it is not clear that their instruments is truly exogenous as peers' parents obesity status or health may be correlated with unobserved variables influencing own BMI. Also, their instruments are *ad hoc* as they are not explicitly derived from the structural form of the model. In our approach, we introduce school fixed effects as well as, for each individual variable, the corresponding contextual variable at the peer level. We can thus identify both endogenous and contextual peer effects. In addition, our instruments are explicitly derived from the structural model.

Using the same dataset, Cohen-Cole and Fletcher (2008b) exploit panel information (wave II in 1996 and wave III in 2001) for adolescents for whom at least one of same-sex friend is also observed over time. Compared with Christakis and Fowler's approach, their analysis introduces time invariant and time dependent environmental variables (at the school level). Friendship selection is controlled for by individual fixed effects. The authors find that peer effects are no longer significant with this specification. As in Trogdon et al. (2008) , their analysis ignores contextual variables, contrary to our approach. Moreover, the friendship network they used in estimations is incomplete, which may underestimate the endogenous peer effect (see Stinebrickner and Stinebrickner, 2006b).

All the studies discussed up to this point focus on peer effects in weight outcomes without analyzing quantitatively the mechanisms by which they may occur. The general issue addressed in this paper is whether the peer effects in weight gain among adolescents partly flow through the *eating habits* channel. This raises in turn two basic issues: a) are there peer effects in fast food consumption?, and b) is there a link between weight gain (or obesity) and fast food consumption? In this paper, we address both issues. The literature on peer effects in eating habits (first issue) is recent and quite limited. In a medical experimental context, Salvy et al. (2008) assess the presence of "peer effect" in pre-adolescent girls' snack intake as a function of the co-eaters' weight status. They show that overweight girls eating with an overweight peer consumed more calories than overweight participants eating with normal weight peers. In a recent natural experiment, Yakusheva et al. (2010) estimate peer effects in explaining weight gain among freshman girls using a similar set up but in school dormitories. Also, they test whether some of the student's weight management behaviours (*i.e.*, eating habits, physical exercise, use of weight loss supplements) can be predicted by her randomly assigned roommate's behaviours. Their results provide evidence of the presence of *negative* peer effects in weight gain. Their results also suggest *positive* peer effects in eating habits, exercise and use of weight loss supplements.

Two caveats of these two studies are their focus on girls and their limited sam-

ple (*e.g.*, recruited participants, freshman level students). Moreover their estimates are likely to underestimate social interactions effects as co-eaters or roommates do not reflect the *true* social network influencing students' weight management behaviours (Stinebrickner and Stinebrickner, 2006b). Finally, these studies do not estimate the causal links between behaviours and weight gain. Our paper finds its basis in this literature as well as the literature on peer effects and obesity discussed above. However, while both works by Salvy et al. (2008) and Yakusheva et al. (2010) rely upon experimental data, we use observational non-experimental data. Thus, peers are not limited to assigned dyads.[12] Rather, they are considered to have social interactions within a school network. This allows for the construction of a social interaction matrix that reflects how social interaction between adolescents in schools occurs in a more realistic setting (as in Trogdon et al., 2008; Renna et al., 2008). An additional originality of our paper lies in the fact that it relies upon a structural (linear-in-means) approach when relating an adolescent's behaviour to that of his peers. Also, the analogy between the forms of the linear-in-means model and the spatial autoregressive (SAR) model allows us to exploit the particularities of this latter model, in particular the natural instruments that are derived from its structural form.

Regarding the second issue, *i.e.*, the relationship between weight gain (or obesity) and fast food consumption, it is an empirical question that is still on the debate table.[13] There is no clear evidence in support of a causal link between fast food consumption and obesity. Nevertheless, most of the literature in epidemiology find evidence of a positive correlation between fast food consumption and obesity (see for a survey, Rosenheck, 2008).[14]

The economic literature reveals to be conservative with respect to this question. It focuses the impact of "exposure" to fast food on obesity. Dunn (2008), using an instrumental variable approach, investigates the relationship between fast food availability and obesity. He finds that an increase in the number of fast food restaurants has a positive effect on the BMI. Similarly, Currie et al. (2010) find evidence that proximity to fast food restaurants has a significant effect on obesity for 9th graders. On the other hand, Chen et al. (2009) found a small but statistically significant effect in favour of a relationship between BMI values and the density of fast food restaurants. Finally, Anderson and Matsa (2011), exploiting the placement of Interstate Highways in rural

[12] Dyads is a common term in social interaction literature. It used to refer to a pair of individuals who have a social relationship

[13] The literature on the impact of physical activity on obesity is also inconclusive. For instance, Berentzen et al. (2008) provide evidence that decreased physical activity in adults does not lead to obesity.

[14] For instance Bowman et al. (2004) finds that children who consumed fast food consumed more total energy.

areas to obtain exogenous variation in the effective price of restaurants, did not find any causal link between restaurant consumption and obesity.

The factors underlying fast food consumption were also investigated. Jeffery and French (1998) show that hours of TV viewing per day and the frequency of meals eaten at fast food restaurants are both positively associated with increase in the BMI of women. One drawback of this study is that it uses a non representative sample (*i.e.*, individuals who volunteered for the study of weight gain prevention). Chou et al. (2005) find a strong positive correlation between exposure to fast food restaurant advertising and the probability that children and adolescents are overweight. This effect seems to be stronger and more significant for girls (Chou et al., 2005). In fact, this influence can be clearly seen as children are more likely to pick up items that are in "Mac Donald's" packaging (Robinson et al., 2007). More generally, Cutler et al. (2003) and Bleich et al. (2008) argue that the increased calorie intake (*i.e.*, eating habits) plays a major role in explaining current obesity rates. Importantly, weight gain prior to adulthood set the stage for weight gain in adulthood.

While most of the economics literature analyses the relationship between adolescents' fast food consumption and their weight gain using an indirect approach (*i.e*, effect to fast food exposure), we adopt a direct approach in this paper. More precisely, we estimate a dynamic model of weight gain as a function of fast food consumption and lagged weight gain. In order to account for the endogeneity of regressors, we follow instrumental methods that were developed in the econometrics literature to estimate panel dynamic models.

2.3 Structural econometric model

In this section, we first propose a linear-in-means peer effects model of the adolescent's fast food consumption (first equation) and discuss the econometric methods we use to estimate it. We then present our dynamic weight gain production function which relates the adolescent's BMI level to his fast food consumption (second equation).

2.3.1 A structural model of peer effects in fast food consumption

Suppose that we have a set of N adolescents i that are partitioned in a set of L networks. A network is defined as a structure (*e.g.*, school) in which adolescents are potentially tied by a friendship link. Each adolescent i in his network has a set of nominated friends N_i of size n_i that constitute his reference group (or peers). We assume that i is excluded from his reference group. Since peers are defined as nominated friends, the number of peers will not be the same for every network member. Let $\mathbf{G}_l$ $(l = 1, \ldots, L)$ be the social interaction matrix for a network l. Its element g_{lij} takes a value of $\frac{1}{n_i}$ when i is friend with j, and zero otherwise.[15] We define y_{li} as the fast food consumed by adolescent i in network l, x_{li} represents the adolescent i's observable characteristics, $\mathbf{y}_l$ the vector of fast food consumption in network l, and $\mathbf{x}_l$ is the corresponding vector for individual characteristics. To simplify our presentation, we look at only one characteristic (*e.g.*, adolescent's mother education).[16] The correlated effects are captured through network fixed effects (the α_l's). They take into account unobserved factors such as preferences of school, school nutrition policies, or presence of fast food restaurants around the school. The ε_{li}'s are the idiosyncratic error terms. They capture i's unobservable characteristics that are not invariant within the network. Formally, one can write the linear-in-means model for adolescent i as follows:

$$y_{li} = \alpha_l + \beta \frac{\sum_{j \in N_i} y_{lj}}{n_i} + \gamma x_{li} + \delta \frac{\sum_{j \in N_i} x_{lj}}{n_i} + \varepsilon_{li}, \tag{2.1}$$

where $\frac{\sum_{j \in N_i} y_{lj}}{n_i}$ and $\frac{\sum_{j \in N_i} x_{lj}}{n_i}$ are respectively his peers' mean fast food consumed and characteristics.[17] In the context of our paper, β is the *endogenous peer effect.* It reflects how the adolescent's consumption of fast food is affected by his peers' mean fast food consumption. It is standard to assume that $|\beta| < 1$. The *contextual peer effect* is represented by the parameter δ. It captures the impact of his peers' mean characteristic on his fast food consumption. It is important to note that the matrices of $\mathbf{G}_l$'s and the vectors of $\mathbf{x}_l$'s are stochastic but assumed strictly exogenous conditional on α_l, that is, $\mathbb{E}(\varepsilon_{li}|\mathbf{x}_l, \mathbf{G}_l, \alpha_l) = 0$. This assumption is flexible enough to allow for correlation between the network's unobserved common characteristics (*e.g.*, school's cafeteria quality) and observed characteristics (*e.g.*, mother's education).[18] Nevertheless, once we condition

[15] Therefore, the $\mathbf{G}_l$ matrix is row normalized.

[16] Later on, in section 3.1.1, we will generalize the model to account for many characteristics.

[17] This structural model can be derived from a choice-theoretic approach where each adolescent's fast food consumption is obtained from the maximization of his quadratic utility function which depends on his individual characteristics, his own fast food consumption and his reference group's mean fast food consumption and mean characteristics. This approach also assumes that social interactions have reached a noncooperative (Nash) equilibrium (see Blume et al., 2010).

[18] In this case $\mathbb{E}(\alpha_l|\mathbf{G}_l, \mathbf{x}_l) \neq 0$.

on these common characteristics, mother's education is assumed to be independent of i's idiosyncratic unobserved characteristics. Let $\mathbf{I}_l$ be the identity matrix for a network l and $\boldsymbol{\iota}_l$ the corresponding vector of ones, the structural model (2.1) for network l can be rewritten in matrix notation as follows:

$$\mathbf{y}_l = \alpha_l \boldsymbol{\iota}_l + \beta \mathbf{G}_l \mathbf{y}_l + \gamma \mathbf{x}_l + \delta \mathbf{G}_l \mathbf{x}_l + \boldsymbol{\varepsilon}_l, \quad \text{for } l = 1, ..., L. \tag{2.2}$$

Note that model (2.2) is similar to a SAR model (*e.g.*, Cliff and Ord, 1981) generalized to allow for contextual and fixed effects (hereinafter referred to as the GSAR model). Since $|\beta| < 1$, $(\mathbf{I}_l - \beta \mathbf{G}_l)$ is invertible. Therefore, in matrix notation, the reduced form of the model can be written as:

$$\mathbf{y}_l = \alpha_l/(1-\beta)\boldsymbol{\iota}_l + (\mathbf{I}_l - \beta \mathbf{G}_l)^{-1}(\gamma \mathbf{I}_l + \delta \mathbf{G}_l)\mathbf{x}_l + (\mathbf{I}_l - \beta \mathbf{G}_l)^{-1}\, \boldsymbol{\varepsilon}_l, \tag{2.3}$$

where we use the result that $(\mathbf{I} - \beta \mathbf{G}_l)^{-1} = \sum_{k=0}^{\infty} \beta^k \mathbf{G}_l^k$, so that the vector of intercepts is $\alpha_l/(1-\beta)\boldsymbol{\iota}_l$, asuming no isolated adolescents.[19]

Equation (2.3) allows us to evaluate the impact of a marginal shock in α_l (*i.e.*, a common exogenous change in fast food consumption within the network) on an adolescent i's fast food consumption, when the endogenous peer effect is taken into account. One has $\partial(E(y_{li}|\cdot)/\partial\alpha_l = 1/(1-\beta)$. This expression is defined as the social multiplier in our model. When $\beta > 0$ (*strategic complementarities* in fast food consumption), the social multiplier is larger than 1. In this case, the impact of the shock is amplified by social interactions.

We then perform a panel-like *within* transformation to the model. More precisely, we average equation (2.3) over all students in network l and subtract it from i's equation. This transformation allows us to address problems that arise from the fact adolescents are sharing the same environment or preferences. Let $\mathbf{K}_l = \mathbf{I}_l - \mathbf{H}_l$ be the matrix that obtains the deviation from network l mean with $\mathbf{H}_l = \frac{1}{n_l}(\iota_l \iota_l')$. The network within transformation will eliminate the correlated effect α_l. Pre-multiplying (2.3) by $\mathbf{K}_l$ yields the reduced form of the model for network l, in deviation:

$$\mathbf{K}_l \mathbf{y}_l = \mathbf{K}_l(\mathbf{I}_l - \beta \mathbf{G}_l)^{-1}(\gamma \mathbf{I}_l + \delta \mathbf{G}_l)\mathbf{x}_l + \mathbf{K}_l(\mathbf{I}_l - \beta \mathbf{G}_l)^{-1}\boldsymbol{\varepsilon}_l. \tag{2.4}$$

[19] When an adolescent is isolated, that is, with an empty group of friends, his intercept is α_l.

Identification

Our structural model raises two basic identification problems.

- Simultaneity

Simultaneity between individual and peer behaviour, also called the *reflection problem* by Manski (1993), may prevent separating contextual effects from endogenous effects. This problem has been analyzed by Bramoullé et al. (2009) when individuals interact through social networks. Let us define $\mathbf{G}$ the block-diagonal matrix with the $\mathbf{G}_l$'s on its diagonal. Assume first the absence of fixed network effects (*i.e.*, $\alpha_l = \alpha$ for all l). In this case, Bramoullé et al. (2009) show that the structural parameters of the model (2.2) are identified if the matrices $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ are linearly independent. This condition is satisfied there are at least two adolescents who are separated by a link of distance 2 within a network. This means that they are not friends but have a common friend (intransitive triad).The intuition is that this provides exclusion restrictions in the model. More precisely, the friends' friends mean characteristics can serve as instruments for the mean friends' fast food consumption. Of course, when fixed network effects are allowed, the identification conditions are more restrictive. Bramoullé et al. (2009) show that, in this case, the structural parameters are if identified if the matrices $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ *and* $\mathbf{G}^3$ are linearly independent. This condition is satisfied when at least two adolescents are separated by a link of distance 3 within a network, i.e., we can find two adolescents who are not friends but are linked by two friends. In this case, $g^3_{lij} > 0$ while $g^2_{ij} = g_{ij} = 0$. Hence, no linear relation of the form $\mathbf{G}^3 = \lambda_0\mathbf{I} + \lambda_1\mathbf{G} + \lambda_2\mathbf{G}^2$ can exist. This condition holds in most friendship networks and, in particular, in the data we use.[20]

- Endogenous network formation

Since adolescents are not randomly assigned into schools, endogenous self-selection through networks may be the source of potentially serious biases in estimating (endogenous + contextual) peer effects. Indeed, if the variables that drive this process of selection are not fully observable, correlations between unobserved network-specific factors and the regressors are potentially important sources of bias. In our approach, we assume that network fixed effects capture these factors. This is consistent with two-step models of link formation. Each adolescent joins a school in a first step, and forms friendship links with others in his school in a second step. In the first step, adolescents self-select into different schools with selection bias due to specific school characteristics. In a second step, link formation takes place within schools randomly or based on observable individual characteristics only. As shown above, network fixed effects are

[20]Identification fails, however, for a number of non trivial networks. This is notably the case for *complete bipartite networks*. In these graphs, the population of students is divided in two groups such that all students in one group are friends with all students in the other group, and there is no friendship links within groups. These include star networks, where one student, at the centre, is friend with all other students, who are all friends only with him.

cancelled out through a panel-like within transformation.

Other types of correlated effects can occur for reasons other than common network factors. For instance, one can think of other uncommon unobserved reasons for which some people might group together. To account for this possibility we allow for error terms to be autocorrelated within networks so that our model structure becomes analogous to that of a generalized spatial autoregressive model with network autoregressive disturbances (hereinafter referred to as the GSARAR model). In this case, the error terms in (2.2) can be written as:

$$\boldsymbol{\varepsilon}_l = \rho \mathbf{G}_l \boldsymbol{\varepsilon}_l + \boldsymbol{\xi}_l, \tag{2.5}$$

where the innovations, $\boldsymbol{\xi}_l$, are assumed to be $i.i.d.(0, \sigma^2 \mathbf{I}_l)$ and $|\rho| < 1$. Given these assumptions, we can write:

$$\boldsymbol{\varepsilon}_l = (\mathbf{I}_l - \rho \mathbf{G}_l)^{-1} \boldsymbol{\xi}_l. \tag{2.6}$$

Allowing for many characteristics and performing a Cochrane-Orcutt-like transformation on the structural model in deviation, the latter is given by the following structural form:

$$\mathbf{K}_l \mathbf{M}_l \mathbf{y}_l = \beta \mathbf{K}_l \mathbf{M}_l \mathbf{G}_l \mathbf{y}_l + \mathbf{K}_l \mathbf{M}_l \mathbf{X}_l \boldsymbol{\gamma} + \mathbf{K}_l \mathbf{M}_l \mathbf{G}_l \mathbf{X}_l \boldsymbol{\delta} + \boldsymbol{\nu}_l, \tag{2.7}$$

where $\mathbf{X}_l$ is the matrix of adolescents' characteristics in the lth network, $\mathbf{M}_l = (\mathbf{I} - \rho \mathbf{G}_l)$ and $\boldsymbol{\nu}_l = \mathbf{K}_l \boldsymbol{\xi}_l$.

The elimination of fixed network effects using a *within* transformation leads to a singular variance matrix such that $E(\boldsymbol{\nu}_l \boldsymbol{\nu}_l' \mid \mathbf{X}_l, \mathbf{G}_l) = \mathbf{K}_l \mathbf{K}_l' \sigma^2 = \mathbf{K}_l \sigma^2$. To resolve this problem of linear dependency between observations, we follow a suggestion by Lee et al. (2010) and applied by Lin (2010). Let $[\mathbf{Q}_l \quad \mathbf{C}_l]$ be the orthonormal matrix of $\mathbf{K}_l$, where $\mathbf{Q}_l$ corresponds to the eigenvalues of 1 and $\mathbf{C}_l$ to the eigenvalues of 0. The matrix $\mathbf{Q}_l$ has the following properties: $\mathbf{Q}_l' \mathbf{Q}_l = \mathbf{I}_{n_l^*}$, $\mathbf{Q}_l \mathbf{Q}_l' = \mathbf{K}_l$ and $\mathbf{Q}_l' \iota = 0$, where $n_l^* = n_l - 1$ with n_l being the number of adolescents in the lth network. Pre-multiplying (2.7) by $\mathbf{Q}_l'$, the structural model can now be written as follows:

$$\mathbf{M}_l^* \mathbf{y}_l^* = \beta \mathbf{M}_l^* \mathbf{G}_l^* \mathbf{y}_l^* + \mathbf{M}_l^* \mathbf{X}_l^* \gamma + \mathbf{M}_l^* \mathbf{G}_l^* \mathbf{X}_l^* \delta + \boldsymbol{\nu}_l^*, \tag{2.8}$$

where $\mathbf{M}_l^* = \mathbf{Q}_l' \mathbf{M}_l \mathbf{Q}_l$, $\mathbf{y}_l^* = \mathbf{Q}_l' \mathbf{y}_l$, $\mathbf{G}_l^* = \mathbf{Q}_l' \mathbf{G}_l \mathbf{Q}_l$, $\mathbf{X}_l^* = \mathbf{Q}_l' \mathbf{X}_l$, and $\boldsymbol{\nu}_l^* = \mathbf{Q}_l' \boldsymbol{\xi}_l$. With this transformation, our problem of dependency between the observations is solved, since we have $E(\boldsymbol{\nu}_l^* \boldsymbol{\nu}_l^{*\prime} \mid \mathbf{X}_l, \mathbf{G}_l) = \sigma^2 \mathbf{I}_{n_l^*}$.

Following Lee et al. (2010), we propose two approaches to estimate the peer effects model (2.8): a maximum likelihood approach (ML) and a generalized spatial two stage least squares (GS-2SLS) approach. The ML approach imposes more structure (normality) than GS-2SLS. Therefore, under some regularity conditions, ML estimators are asymptotically more efficient than GS-2SLS ones when the restrictions it imposes are valid.

Maximum Likelihood (ML)

Assuming that $\boldsymbol{\nu}_l^*$ is a n_l^*-dimensional normally distributed disturbance vector, the log-likelihood function is given by [21]:

$$\ln \mathbb{L} = \frac{-n^*}{2}\ln(2\pi\sigma^2) + \sum_{l=1}^{L}\ln|\mathbf{I}_{n_l^*} - \beta\mathbf{G}_l^*| + \sum_{l=1}^{L}\ln|\mathbf{I}_{n_l^*} - \rho\mathbf{M}_l^*| - \frac{1}{2\sigma^2}\sum_{l=1}^{L}\boldsymbol{\nu}_l^{*\prime}\boldsymbol{\nu}_l^*, \quad (2.9)$$

where $n^* = \sum_{l=1}^{L} n_l^* = N - L$, and, from (2.8), $\boldsymbol{\nu}_l^* = \mathbf{M}_l^*(\mathbf{y}_l^* - \beta\mathbf{G}_l^*\mathbf{y}_l^* - \mathbf{X}_l^*\gamma - \mathbf{G}_l^*\mathbf{X}_l^*\delta)$. Maximizing (2.9) with respect to $(\beta, \boldsymbol{\gamma}', \boldsymbol{\delta}', \rho, \sigma)$ yields the maximum likelihood estimators of the model.[22] Interestingly, the ML method is implemented after the elimination of the network fixed effects. Therefore, the estimators are not subject to the incidental parameters problem that may arise if the number of fixed effects increases with the the size of the networks sample.

Generalized spatial two stage least squares (GS-2SLS)

To estimate the model (2.8), we also adopt a generalized spatial two-stage least squares procedure presented in Lee et al. (2010). This approach provides a simple and tractable numerical method to obtain asymptotically efficient IV estimators within the class of IV estimators. In the case of our paper this method will consist of a two-step estimation.[23] To simplify the notation, let $\mathbf{X}^*$ be a block-diagonal matrix with $\mathbf{X}^*{}_l$ on its diagonal, $\mathbf{G}^*$ be a block-diagonal matrix with $\mathbf{G}^*{}_l$ on its diagonal, and $\mathbf{y}^*$ the concatenated vector of the y_l^*'s over all networks.

[21] It is important to note that we do not account for the count nature of the dependant variable. We expect that this is bias our estimates downwards.

[22] For computational simplicity, one can concentrate the log-likelihood function (2.9) and maximize the concentrated log-likelihood function. See Lee et al. (2010) for more details.

[23] Note that for this particular case we impose $\rho = 0$ and thus $M_l = I_l$.

Now, let us denote by $\tilde{\mathbf{X}}^*$ the matrix of explanatory variables such that $\tilde{\mathbf{X}}^* = [\mathbf{G}^*\mathbf{y}^* \quad \mathbf{X}^* \quad \mathbf{G}^*\mathbf{X}^*]$. Let $\mathbf{P}$ be the weighting matrix such that $\mathbf{P} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$, and $\mathbf{S}$ a matrix of instruments such that $\mathbf{S} = [\mathbf{X}^* \quad \mathbf{G}^*\mathbf{X}^* \quad \mathbf{G}^{*2}\mathbf{X}^*]$. In the first step, we estimate the following 2SLS estimator:

$$\hat{\boldsymbol{\theta}}_1 = (\tilde{\mathbf{X}}^{*\prime}\mathbf{P}\tilde{\mathbf{X}}^*)^{-1}\tilde{\mathbf{X}}^{*\prime}\mathbf{P}\mathbf{y}^*,$$

where $\hat{\boldsymbol{\theta}}_1$ is the first-step 2SLS vector of estimated parameters $(\hat{\boldsymbol{\gamma}}_1', \hat{\boldsymbol{\delta}}_1', \hat{\beta}_1)$ of the structural model. This estimator is consistent but not asymptotically efficient within the class of IV estimators.

Now, in the second step, we estimate a 2SLS using a new matrix of instruments $\hat{\mathbf{Z}}$ given by:

$$\hat{\mathbf{Z}} = [\mathbf{G}^*\hat{\mathbf{y}}^* \quad \mathbf{X}^* \quad \mathbf{G}^*\mathbf{X}^*],$$

where $\mathbf{G}^*\hat{\mathbf{y}}^*$ is computed from the first-step 2SLS reduced form (pre-multiplied by $\mathbf{G}^*$):

$$\mathbf{G}^*\hat{\mathbf{y}}^* = \mathbf{G}^*(\mathbf{I} - \hat{\beta}_1\mathbf{G}^*)^{-1}(\mathbf{X}^*\hat{\boldsymbol{\gamma}}_1 + \mathbf{G}^*\mathbf{X}^*\hat{\boldsymbol{\delta}}_1).$$

We then estimate:

$$\hat{\boldsymbol{\theta}}_2 = (\hat{\mathbf{Z}}'\tilde{\mathbf{X}}^*)^{-1}\hat{\mathbf{Z}}\mathbf{y}^*.$$

This estimator can be shown to be consistent and asymptotically best IV estimator. Its asymptotic variance matrix is given by $N[\mathbf{Z}'\tilde{\mathbf{X}}^*\mathbf{R}^{-1}\tilde{\mathbf{X}}^*\mathbf{Z}]^{-1}$. The matrix $\mathbf{R}$ is consistently estimated by $\hat{\mathbf{R}} = s^2\frac{\hat{\mathbf{Z}}'\hat{\mathbf{Z}}}{N}$, where $s^2 = N^{-1}\sum_{i=1}^{N}\hat{u}_i{}^2$ and $\hat{u}_i$ are the residuals from the second step. It is important to note that, as in Kelejian and Prucha (1998), we assume that errors are homoscedastic. The estimation theory developed by Kelejian and Prucha (1998) under the assumption of homoscedastic errors does not apply if we assume heteroscedastic errors (Kelejian and Prucha, 2010).

2.3.2 A weight gain production function

In this section, we propose a weight gain production function that relates an adolescent's BMI in time t to his lagged BMI, his fast food consumption as well as his own characteristics in period t. Let y_{it}^b be an adolescent i's BMI level at time t, and y_{it}^f be the adolescent's fast food consumption. Then, for a given vector of characteristics $\tilde{\mathbf{x}}_{it}$, the weight gain production function can be formally expressed as follows (for notational simplicity we suppress l):

$$y_{it}^b = \pi_0 + \pi_1 y_{i,t-1}^b + \pi_2 y_{it}^f + \boldsymbol{\pi}_3\tilde{\mathbf{x}}_{it} + \eta_{it}, \tag{2.10}$$

where

$$\eta_{it} = \mu_i + \zeta_{it},$$

with μ_i representing the individual i's time-invariant error component (unobserved heterogeneity) and ζ_{it}, his idiosyncratic error that may change across t. We consider that μ_i is a fixed effect, $|\pi_1| < 1$ and the error ζ_{it} is serially uncorrelated. (A test of this latter assumption is provided in the empirical section). As discussed earlier, our interest in this production function goes beyond a mere association between fast food consumption and weight gain. We are particularly interested to analyze the magnitude of a change in BMI resulting from a common exogenous shock on fast food consumption within the network, when peer effects are taken into account. Our two equation model allows us to compute this result. Partially differentiating (2.10) with respect to y^f_{it} and using the social multiplier $[= 1/(1-\beta)]$ yields the magnitude of a short run change in BMI (*i.e.*, for $y^b_{i,t-1}$ given) resulting from a common marginal shock on fast food consumption: $\partial E(y^b_{it}|\cdot)/\partial \alpha_l = \frac{\pi_2}{1-\beta}$. This expression entails two components: the impact of the fast food consumption on the BMI $(= \pi_2)$ and the multiplier effect $(= \frac{1}{1-\beta})$.[24]

At this point it is important to mention that OLS estimates of (2.10) will not be consistent for two reasons. First, the adolescent's fast food consumption is not exogenously determined and may be affected by his own BMI. In addition, there may exist a correlation between lagged dependent variable and the error term due to the presence of a time-invariant error component. One way to resolve this issue is to apply a first difference. While such a transformation wipes out all individual time-invariant characteristics, it has the advantage of making the correlation between right hand side regressors and the error term easier to handle. Formally, assuming that all characteristics (except age) are time-invariant, the transformed model can be written as follows:

$$\Delta y^b_{it} = \tilde{\pi_0} + \pi_1 \Delta y^b_{i,t-1} + \pi_2 \Delta y^f_{it} + \Delta \zeta_{it}, \tag{2.11}$$

where Δ is the first difference operator. To resolve the problem of correlation between the right hand side variables and the error term one can instrument for $\Delta y^b_{i,t-1}$ and Δy^f_{it}. To instrument the lagged dependent variable we can either use $\Delta y^b_{i,t-2}$ or $y^b_{i,t-2}$ (Hsiao, 1981). Following the suggestion of Arellano and Bond (1991), we instrument it using y^b_{t-2}. We also use birth weight x_{bw}. High birth weight is associated with overweight in adolescence (Gillman et al., 2003; Sorensen et al., 1997). It is therefore expected that it would affect y^b_{it} only through it impact on $y^b_{i,t-1}$. As for the fast food consumption variable Δy^f_{it}, valid instruments could be $y^f_{i1}, y^f_{i2}, y^f_{i3}, \ldots, y^f_{i,s-1}$ for an equation differenced at $t = s$. Unfortunately such information is not available in our dataset. We thus instrument it using birth weight and the strictly exogenous variables that we have used in our peer effects model.

[24] We assume that the individuals' weight gain is symmetric over all weight categories.

2.4 Data and Descriptive Statistics

The Add Health survey is a longitudinal study that is nationally representative of American adolescents in grades 7 through 12. It is one of the most comprehensive health surveys that contains fairly exhaustive social, economic, psychological and physical well-being variables along with contextual data on the family, neighbourhood, community, school, friendships, peer groups, romantic relationships, *etc.* In the first wave (September 1994 to April 1995), all students (around 90 000) attending the randomly selected high schools were asked to answer a short questionnaire. An in-home sample (core sample) of approximately 20 000 students was then randomly drawn from each school. These adolescents were asked to participate in a more extensive questionnaire where detailed questions were asked. Information on (but not limited to) health, nutrition, expectations, parents' health, parent-adolescent relationship and friends nomination was gathered.[25] This cohort was then followed in-home in the subsequent waves in 1996 (wave II) and 2001 (wave III). The extensive questionnaire was also used to construct the saturation sample that focuses on 16 selected schools (about 3000 students). Every student attending these selected schools answered the detailed questionnaire. There are two large schools and 14 other small schools. All schools are racially mixed and are located in major metropolitan areas except one large school that has a high concentration of white adolescents and is located in a rural area. Consequently, fast food consumption may be subject to downward bias if one accepts the argument that the fast food consumption among white adolescents is usually lower than that of black adolescents.[26]

In this paper we use the saturation sample of wave II in-home survey to investigate the presence of peer effects in fast food consumption.[27] One of the innovative aspects of this wave is the introduction of the nutrition section. It reports among other things food consumption variables (*e.g.*, fast food, soft drinks, desserts, *etc.*).[28] This allows us to depict food consumption patterns of each adolescent and relate it to that of his peer group. In addition, the availability of friend nomination allows us to trace school friends and thus construct friendship networks.[29] To estimate the weight gain production function, we considered information from wave I, wave II and wave III.

[25] Adolescents were asked to nominate either 1 female friend and 1 male friend or 5 female friends and 5 male friends.Friendship nomination is thus top coded this may introduce some bias (see Chandrasekhar and Lewis, 2011)

[26] It is unclear whether the black population consumes more fast food than the white population. Block et al. (2004) provides evidence that predominantly black neighbourhoods have relatively more fast food restaurants than predominantly white neighbourhoods.

[27] It includes all meals that are consumed at a fast food restaurant such as McDonald's, Burger King, Pizza Hut, Tacco Bell.

[28] Frequency of fast food consumption is reported in terms of day visits, it does not capture the frequency per day.

[29] In the saturation sample, all students are asked to name their friends.

We exploit friends nominations to construct the network of friends. Thus, we consider all nominated friends as network members regardless of the reciprocity of the nomination. If an adolescent nominates a friend but that this friend does not nominate him, then only one link is assigned (i.e., a link from that adolescent to the nominated friend and no link is assigned from the nominated friend to him).[30]

2.4.1 Descriptive statistics

In our peer effects model, the dependent variable of interest is fast food consumption, as approximated by the reported frequency (in days) of fast food restaurant visits in the past 7days. Table 1 reports respectively the mean and the standard deviation of the endogenous variable, the covariates used and other relevant characteristics.[31] We note that on average, adolescents' fast food consumption is around 2.33 times/week. This is consistent with the frequency reported by the Economic Research Service of the United States Department of Agriculture (Lin et al., 1996). Around 62% of the adolescents consumed fast food twice or more in the past week and 44% of the adolescents who had consumed fast food did so 3 times in the past week. We also inspected the data to check for the presence of parental intervention in eating habits. We note that parents are absent 38% of the times during the consumption of evening meals and that 83 % of children have the freedom to chose their meals.[32] Evening meals are the only time parents have the power to check on the quality/quantity of the food consumed. This coupled with the freedom to choose food leaves the door wide open for peer influence in fast food consumption.

The covariates of the fast food peer effect equation include the adolescent's personal characteristics, family characteristics as well as the corresponding contextual social effects.The personal characteristics are gender, age, ethnicity (white or other) and grade. We observe that 50% of the sample are females, that the mean age is 16.3 years and that 57% are white. Family characteristics are dummies for mother and father education. We observe that around 45% of mothers and fathers have at least some college education. To control further for parents' income we use child allowance as a proxy. An adolescent's allowance is on average 8.28 $ per week, around 50% of the adolescents in our the sample have a weekly allowance. At this point, it is important to highlight that since we use cross section data, we do not have to control for fast food prices as they are taken into account by network fixed effects. As for the weight gain production function, the dependent variable that we use is the variation in the BMI between waves two and

[30] It is important to note that we do not impose symmetry on the social interaction matrix.

[31] Only 15% of the students do not consume fast food at all.

[32] These figures are not reported in the paper but can be provided upon request.

three. The covariates are the fast food consumption, the lagged BMI (Table 2.2), age, and all time-invariant variables appearing in Table 2.1.

2.4.2 The Construction of the Graph Matrix

We construct a sub-matrix of graph for each school separately (matrice $\mathbf{G}_l$) and then we include all these sub-matrices in the block-diagonal matrix $\mathbf{G}$.[33] As we have no prior information about how social interaction takes place, we assume, as in most studies, that an adolescent is equally influenced by his nominated friends. Further, we assume this influence decreases with the number of friends. In each school we eliminate adolescents for which we have missing values. We allow the sub-matrices to contain adolescents who are isolated. Since these latter may be friends with other adolescents in the network, they may affect the network even if they claim not to have any friends at all. They also introduce variability that helps the identification of the model. We also do not impose symmetry on the $\mathbf{G}$ matrix. This means that we allow for non-reciprocated friendships. This imposes less restriction on the social interaction and mimics better the social interaction.[34]

As mentioned earlier, Bramoullé et al. (2009) show that, the structural parameters are identified if the matrices $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ *and* $\mathbf{G}^3$ are linearly independent. One way check whether these four matrices are linearly independent as follows. First, vectorize each matrix, that is, stack its columns on top of each other to form a matrix L. Second, verify whether the matrix formed by concatenating these stacked vectors has rank four. An alternative more general diagnosis check for linear independence is to do the first step as above and compute $L'L$ and scale it into a correlation form. Then compute the Belsley, Kuh, and Welsch *condition index* by finding the square root of the ratio of the largest eigen value to the smallest eigen value. If this ratio is below 30, then collinearity is said not to be a problem and linear independance of the four matrices is verified. In our data, the reflection problem is solved since $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ and $\mathbf{G}^3$ are linearly independent and the condition index value is 2.21.

[33] Following the previous literature and given the lack of information on this matter, we assume that there can be social interactions within each school but no interactions across schools.

[34] In this paper we do not account for partial observability of the social network (five female friends and five male friends)

2.5 Results

2.5.1 Baseline: *OLS* peer effects estimates

We first estimate a naive *OLS* of the peer effects model where we regress the fast food consumption of an adolescent on the average fast food consumption of his peers, his individual characteristics as well as the average characteristics of his peers. We then apply a panel-like *within* transformation to account for correlated effects (OLS_w). It is clear that the estimates of naive *OLS* and OLS_w are inconsistent. The former ignores both correlated effects and simultaneity problems while the latter ignores simultaneity problems. However, they are reported to provide a baseline for this study.

Estimation results reported in Table 2.3 show that there is a positive significant peer influence in fast food consumption. According to the naive *OLS* estimates, an adolescent would increase his weekly frequency (in days) of fast food restaurant visits by 0.21 in response to an extra day of fast food restaurant visits by his friends. On average, this corresponds to an increase of 9% (= 0.21/2.33). OLS_w estimate is slightly lower (= 0.15, or 6.6%). This reduction in the estimated effect may partly be explained by the fact that adolescents in the same reference group tend to choose a similar level of fast food consumption partly because they are alike or face a common environment. How can we compare these results to those obtained previously in the related literature? Although there are few studies that investigated the presence of peer effects in fast food consumption, a richer body of literature has investigated a tangent issue : obesity. In their paper Trogdon et al. (2008) show *OLS* results for peer effects in obesity of 0.30. Also, Renna et al. (2008) reports endogenous effects of 0.16 for *OLS* estimates. This makes our *OLS* estimates comparable to those obtained in the literature on obesity.

As for the individual characteristics they seem to be increasing in age, father education and weekly allowance. Turning our attention to the contextual peer effects, we notice that fast food consumption decreases with mean peers' mother's education and increases with mean peers' father's education. The former result indicates that friends' mother education negatively affects an adolescent's fast food consumption.

2.5.2 ML and GS-2SLS peer effects estimates

Next, we estimate our linear-in-means (or GSAR) model with school fixed effects and using ML. We then estimate a more general version of this model by allowing network

autoregressive disturbances (GSARAR model). Also, given that ML approach imposes normality on the error term, we relax this assumption and estimate the model using a distribution free approach : GS-2SLS.

Estimation results displayed in Table 2.4 show a positive and statistically significant endogenous effect of 0.13 (or 5.5%) for the GSAR model. This effect is slightly smaller than the ones obtained in the previous section. However, based on the more general GSARAR model, the estimated autocorrelation coefficient is significant and negative. Since it is significant, this indicates that the GSAR model is rejected as a particular case of the GSARAR model. Also, the fact that it is negative may provide evidence that friendship might be for other purposes than having a common preference in the formation of friends for fast food consumption.[35]

With the GSARAR specification, the endogenous peer effect remains statistically significant but increases to 0.37, suggesting that an adolescent would increase his weekly frequency (in days) of fast food restaurant visits by 0.37 (or 15.9%) in response to an extra day per week of fast food restaurant visits by his friends. The social multiplier associated with an exogenous increase in an adolescent fast food consumption is 1.59 ($= \frac{1}{1-0.37}$), which reflects a fairly strong endogenous peer effect.[36] The increase in the endogenous effect coefficient as a result of a GSARAR specification on the error term is comparable to the one obtained by Lin (2010) in an empirical application of such a model on peer effects in academic achievement. As for the magnitude of the endogenous effects, it remains lower than the ones obtained in the literature on peer effects in obesity.[37] Recall that Trogdon et al. (2008) estimate for the endogenous effect is 0.52 using an instrumental approach.

When we relax the normality assumption, the endogenous effect resulting from GS-2SLS estimation turns out to be smaller than the one obtained by ML (0.11 instead of 0.13). However, it is no longer significant. This does not come as a surprise, as less structure is not without a cost in precision. This makes the GSARAR specification preferable to others, at least as long as the normality assumption is an appropriate assumption. To sum up, we can say that results in general are consistent with the hypothesis that fast food consumption is linked to issues of identity and friends (Story et al., 2002).

[35] Lin (2010) also obtains a negative estimated autocorrelation parameter when her model takes endogenous and contextual effects into account.

[36] As suggested by Glaeser et al. (2003), large social multipliers tend to occur when the endogenous effect is 0.33 or more.

[37] One possible explanation is that we are estimating peer effects using one potential behavioural channel.

As for individual effects, they follow fairly the baseline model. Fast food consumption is positively associated with age and father's education as well as positively associated with weekly allowance.[38] Mother's education seems to have a negative but insignificant impact on fast food consumption. It is important to note that while the general perception is that fast food is an *inferior* good, the empirical evidence suggests that there is a positive income elasticity (McCracken and Brandt, 1987; Jekanowski et al., 2001; Aguiar and Hurst, 2005). Thus, the positive relation between fast food consumption and allowance is therefore in line with the positive relation between income and fast food consumption.

Turning our attention to the contextual social effects, fast food consumption increases with mean peers' father's education and decreases with mean peers' mother's education. This suggests that adolescents are perhaps more influenced by their friends' mothers than their own.

2.5.3 Weight gain production function estimates

Estimation results presented in the earlier sections are consistent with the presence of peer effects in fast food consumption. Nevertheless, we still need to provide evidence of the presence of a relationship between fast food consumption and weight gain. In this section we report estimates of the weight gain production function presented earlier. As noted above, the variables we used to instrument the fast food consumption variable Δy_{it}^{f} are the instruments previously used in the fast food consumption model.

Results from the Arellano and Bond (1991) estimator are reported in Table 2.5. Note that since the model is estimated using a first differencing (fixed effects) approach, parameters associated with time-invariant covariates are not identified and therefore do not appear in the table (the constant is an estimate of age parameter). In line with our expectations, results reveal a positive significant impact of a change in fast food consumption on the BMI level. An extra day of fast food restaurant visits per week increases weight by 0.36 BMI points (or by 1.54%) within a year.[39]

The presence of a causal link between fast food consumption and BMI does not come as a surprise since previous findings have been pointing in this direction (Levitsky et al.,

[38] McLellan et al. (1999) found results pointing in a similar direction and suggest that limiting pocket money may be a good way to promote healthy adolescent behaviour.

[39] We investigated the possibility of an omitted variable bias by introducing three (endogenous) physical exercise variables from Add Health in the weight gain equation: exercising activities, active sports, and recreational sports. None of these variables were significant and the fast food estimated effect (= 0.34), while a little smaller, was quite robust to this modification.

2004; Niemeier et al., 2006; Rosenheck, 2008). Somewhat surprisingly, lagged BMI level has a negative effect on current BMI level ($= -0.727$). This suggests that an exogenous shock on weight gain has a stronger effect on BMI in the short term than in the long term. This may partly be explained by the fact that given a past increase in his BMI an adolescent may be induced to adopt more healthy eating habits.

The consistency of our estimates rely heavily on the validity of the instruments that are used. In order to test the validity of the instruments a Sargan test is computed. The test statistic reported in Table 2.5 indicates that we do not reject the joint null hypothesis that the instruments are valid and that the over-identification restrictions are satisfied. Observe that the Sargan test provides an (indirect) test for serial autocorrelation of the idiosyncratic error ζ_{it}. Indeed, the variable y^b_{t-2} would not be a valid instrument under serial autocorrelation. Arellano and Bond (1991) suggest to use this test when the length of the panel is too short to perform a direct serial autocorrelation test (*e.g.*, when $T = 3$ as in our case).

Combining the impact of fast food on weight gain with the social multiplier, our results suggest that, within a network, an extra day of fast food restaurant visits per week leads to a BMI increase of 0.57 points ($\frac{0.36}{1-0.37}$), or 2.4% on average, within a year. These results highlight the role of peer effects in fast food consumption as one of the transmission mechanisms through which weight gain is amplified.

2.6 Conclusion

This paper investigates whether peer effects in adolescent weight gain partly flow through the eating habits channel. We first attempt to study the presence of significant endogenous peer effects in fast food consumption. New methods based on spatial econometric analysis are used to identify and estimate our model, under the assumption that individuals interact through a friendship social network. Our results indicate that an increase in his friends' mean fast food consumption induces an adolescent to increase his own fast food consumption. This peer effect amplifies through a social multiplier the impact of any exogenous shock on fast food consumption. Our estimated social multiplier is 1.59.

We also estimate a dynamic weight gain production function which relates the adolescent's Body Mass Index to his fast food consumption. Results are in line with our expectations; they reveal a positive significant impact of a change in fast food consumption on the change in BMI. Specifically, a one-unit increase in the weekly frequency (in

days) of fast food consumption produces an increase in BMI by 1.5% within a year. This effect reaches 2.4% when the social multiplier is taken into account. Coupled with the reduction in the relative price of fast food and the increasing availability of fast food restaurants over time, the social multiplier could exacerbate the prevalence of obesity in the years to come. Conversely, this multiplier may contribute to the decline of the spread of obesity and the decrease in health care costs, as long as it is exploited by policy makers through tax and subsidy reforms encouraging adequate eating habits among adolescents.

There are many possible extensions to this paper. From a policy perspective, it would be interesting to investigate the presence of peer effects in physical activity of adolescents. A recent study by Charness and Gneezy (2009) finds that there is room for intervention in peoples' decisions to perform physical exercise through financial incentives. It would be thus valuable to investigate whether there is a social multiplier that can be exploited to amplify these effects. Furthermore, in the same way, it would be interesting to study the presence of peer effects weight perceptions. So far, most of the peer effects work has focused mainly on outcomes (BMI). At the methodological level, a possible extension would be to relax the normality assumption and to assume a Poisson or a Negative Binomial distribution to account for the count nature of the consumption data at hand. As far as we know, no work has been carried out in this area. Another interesting extension would be to endogenies the social network formation. Christakis et al. (2010) have a recent paper in which they develop a method to endogenies network formation. Also, a recent work by Dinardo and Strange (2010) suggest that the relation between the BMI and health outcomes is U shaped. Given recent development Quantile Instrumental Variable approach to Spatial Auto-regressive models, it would be interesting to investigate this possibility. Finally, it would be most useful to develop a general approach that would allow same sex and opposite sex peer effects to be different for both males and females.

Table 2.1: Descriptive Statistics

Variable	Mean	S.D
Fast Food Consumption[a]	2.33	1.74
Female	.50	.50
Age	16.36	1.44
White	.57	.49
Black	.15	.34
Asian	.01	.09
Native	.13	.33
Other	.14	.35
Mother Present	.85	.35
Mother Education		
No high school degree	.15	.35
High school/GED/Vocational Instead of high school	.36	.48
Some College/Vocational After high school	.21	.39
College	.18	.38
Advanced Degree	.06	.24
Don't Know	.04	.20
Father Education		
No high school degree	.16	.36
High school/GED/Vocational Instead of high school	.33	.47
Some College/Vocational After high school	.17	.37
College	.18	.38
Advanced Degree	.08	.26
Don't Know	.06	.24
Missing	.02	.16
Grade 7-8	.11	.32
Grade 9-10	.27	.44
Grade 11-12	.62	.48
Allowance per week	8.28	11.65
Observations:		2355

[a]Frequency (in days) of fast food restaurant visits in the past week.

Table 2.2: Descriptives for weight indices

Variable	Mean	S.D
Bmi (wave1)	22.81	4.59
Bmi (wave2)	23.24	4.76
Bmi (wave3)	26.78	6.31
Birthweight	7.40	1.38

Table 2.3: Peer effects in fast food consumption OLS and OLS within

	OLS			OLSw		
	Coef.		**S.E**	**Coef.**		**S.E**
Endogenous Peer Effects	0.2078	***	0.0331	0.1548	***	0.0344
Individual Characteristics						
Female	-0.0721		0.0787	-0.0847		0.0789
Age	0.1559	***	0.0434	0.1315	***	0.0461
White	-0.1076		0.0940	-0.0602		0.1127
Mother Present	-0.0152		0.0997	-0.0358		0.0989
Mother No High School (Omitted)						
Mother High School	-0.0848		0.1195	-0.0455		0.1202
Mother Some College	-0.0377		0.1335	-0.0210		0.1340
Mother College	0.0214		0.1421	-0.0137		0.1425
Mother Advanced	-0.0259		0.1875	-0.0353		0.1877
Mother Don't Know	-0.1714		0.2067	-0.2124		0.2059
Father No High School (Omitted)						
Father High School	0.2743	**	0.2067	0.2682	**	0.1167
Father Some College	0.2117		0.2067	0.1971		0.1338
Father College	0.3115	**	0.1375	0.2592	*	0.1381
Father Advanced	0.1732		0.1752	0.1294		0.1760
Father Don't Know	0.2778		0.1756	0.2393		0.1750
Father Missing	0.0908		0.2338	0.0477		0.2331
Grade 7-8 (Omitted)						
Grade 9-10	0.0883		0.1931	-0.0776		0.2183
Grade 11-12	0.3164		0.2265	0.1269		0.2526
Allowance per week	0.0093	***	0.0031	0.0074	**	0.0031

continued on next page

Table 2.3: *Continued*

	OLS			OLSw		
	Coef.		**S.E**	**Coef.**		**S.E**
Contextual Peer Effects						
Female	-0.0898		0.1245	-0.1071		0.1285
Age	-0.0321		0.0215	0.0316		0.0718
White	0.0111		0.1244	-0.0055		0.1694
Mother Present	0.0773		0.1668	0.1008		0.1707
Mother No High School (Omitted)						
Mother High School	-0.3878	**	0.1868	-0.2977		0.1913
Mother Some College	-0.3947	*	0.2127	-0.3825	*	0.2168
Mother College	-0.2531		0.2180	-0.2935		0.2213
Mother Advanced	-0.7011	**	0.3089	-0.5954	*	0.3112
Mother Don't Know	-0.4337		0.3598	-0.4150		0.3610
Father No High School (Omitted)						
Father High School	0.2060		0.1943	0.2999		0.1914
Father Some College	0.3639	*	0.2128	0.3890	*	0.2139
Father College	0.2850		0.2238	0.3068		0.2263
Father Advanced	0.2760		0.2891	0.2171		0.2953
Father Don't Know	0.4737		0.2995	0.5358	*	0.3001
Father Missing	0.6931		0.4619	0.7692	*	0.4640
Grade 7-8 (Omitted)						
Grade 9-10	-0.0769		0.2383	0.0104		0.2773
Grade 11-12	-0.0094		0.2630	-0.0396		0.3388
Allowance per week	0.0056	**	0.0053	0.0043		0.0054
Constant	-0.5199		0.6618			
N=2239						

*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

Table 2.4: Peer effects in fast food consumption GSAR, GSARAR and GS-2SLS

	MLE						GS-2SLS		
	GSAR		**S.E**	**GSARAR**		**S.E**	**GSAR**		**S.E**
Endogenous Peer Effects	0.1292	***	0.0292	0.3656	***	0.0657 *	0.1102		0.3929
ρ				-0.2577	***	0.0732			
Individual Characteristics									
Female	-0.0783		0.0782	-0.0726		0.0780	-0.0838		0.0793
Age	0.1401	***	0.0440	0.1426	***	0.0426	0.1345	**	0.0531
White	-0.0618		0.1111	-0.0511		0.1113	-0.0618		0.1137
Mother Present	-0.0319		0.0986	-0.0334		0.0982	-0.0375		0.1000
Mother No High School (Omitted)									
Mother High School	-0.0329		0.1190	-0.0420		0.1190	-0.0436		0.1214
Mother Some College	-0.0102		0.1327	-0.0075		0.1327	-0.0161		0.1409
Mother College	0.0045		0.1410	0.0084		0.1410	-0.0142		0.1426
Mother Advanced	-0.0156		0.1862	-0.0243		0.1860	-0.0365		0.1880
Mother Don't Know	-0.2190		0.2042	-0.2137		0.2029	-0.2137		0.2062
Father No High School (Omitted)									
Father High School	0.2777	**	0.1157	0.2514	**	0.1156	0.2689	**	0.1169
Father Some College	0.2031		0.1326	0.1735		0.1324	0.1956		0.1344
Father College	0.2777	**	0.1372	0.2646	*	0.1367	0.2577	*	0.1388
Father Advanced	0.1340		0.1746	0.1297		0.1743	0.1275		0.1769
Father Don't Know	0.2514		0.1735	0.2307		0.1725	0.2419		0.1766
Father Missing	0.0547		0.2308	0.0334		0.2291	0.0515		0.2355
Grade 7-8 (Omitted)									
Grade 9-10	-0.1457		0.2138	-0.1398		0.2131	-0.0789		0.2186
Grade 11-12	0.0268		0.2439	0.0374		0.2430	0.1249		0.2533
Allowance per week	0.0076	**	0.0031	0.0074	***	0.0031	0.0075	**	0.0032

continued on next page

Table 2.4: *Continued*

	MLE						GS-2SLS	
	GSAR		**S.E**	**GSARAR**		**S.E**	**GSAR**	**S.E**
Contextual Peer Effects								
Female	-0.1571		0.1255	-0.1290		0.1212	-0.1108	0.1325
Age	-0.0369	*	0.0216	-0.0727	***	0.0221	0.0359	0.0811
White	0.0068		0.1372	0.0245		0.1279	-0.0159	0.1922
Mother Present	0.0623		0.1660	0.0818		0.1621	0.1077	0.1811
Mother No High School (Omitted)								
Mother High School	-0.3200	*	0.1867	-0.2248		0.1833	-0.3000	0.1924
Mother Some College	-0.4276	**	0.2116	-0.3745	*	0.2065	-0.3881	* 0.2224
Mother College	-0.3464		0.2166	-0.3179		0.2130	-0.3080	0.2549
Mother Advanced	-0.6582	**	0.3066	-0.5436	*	0.2971	-0.5774	* 0.3490
Mother Don't Know	-0.4701		0.3560	-0.3063		0.3539	-0.4038	0.3741
Father No High School (Omitted)								
Father High School	0.3177	*	0.1921	0.2887		0.1871	0.3299	0.3219
Father Some College	0.3873	*	0.2103	0.3895	*	0.2060	0.4051	0.2561
Father College	0.3194		0.2223	0.2638		0.2171	0.3298	0.3028
Father Advanced	0.1744		0.2897	0.1365		0.2817	0.2340	0.3305
Father Don't Know	0.5532	*	0.2959	0.4869	*	0.2899	0.5683	0.4140
Father Missing	0.7748	*	0.4606	0.7048		0.4500	0.7769	* 0.4690
Grade 7-8 (Omitted)								
Grade 9-10	0.1923		0.2668	0.1816		0.2549	0.0057	0.2804
Grade 11-12	0.3177		0.2849	0.2531		0.2691	-0.0340	0.3422
Allowance per week	0.0025		0.0053	0.0010		0.0051	0.0048	0.0069
Constant								
N=2239								

*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

Log likelihood for the GSAR and the GSARAR are respectivley -4488.84623 and GSARAR -4486.3837

Table 2.5: Weight gain Production Function, Arellano and Bond

	Coefficient	S.E	
Constant	3.7206	0.13991	***
Δ BMI_{t-1}	-0.72737	0.19108	***
Δ Fast $food_t$	0.35761	0.17936	**
Sargan test Chi2(57)		61.955	
N		1445	

Chapter 3

A Friend a Day may Keep the Doctor Away: Is Physical Activity Contagious?

3.1 Introduction

Sedentary lifestyle (i.e., physical inactivity) imposes a heavy burden on the economy and is an expensive public health problem. It has been associated with higher medical expenditures (for a complete review see McInnes and Shinogle, 2009) and loss in productivity. Whilst the costs and benefits of physical exercise are fairly clear to most, a large proportion of the general population of the United States as well as their adolescents fail to embrace the recommendations of the World Health Organization regarding the minimum required amount of exercising.[1]

Physical activity habits can be analyzed in two broad dimensions: behavioural and environmental (i.e., related to their location). This paper will focus on the behavioural dimension.[2] Physical inactivity is often attributed to self-control issues or procrastination as defined by Akerlof (1991).[3] In the presence of a *salient cost*, failing to exercise is a consequence of an individual's incapacity to save health for the future.[4] For in-

[1] In 2007, for instance only 35% of adolescents met the recommended level of physical activity. The recommended amount of activity is 30 minutes of moderate exercise daily or 20 minutes of vigorous exercise 3 times a week.

[2] According to Kohl III and Hobbs (1998) there are three broad dimensions, the first is psychologic and developmental the second is environmental and the third is social. In this paper the psycho-social dimension is referred to as behavioural.

[3] Procrastination is the consequence of time preferences characterized by hyperbolic discounting.

[4] Also, the present predominant sedentary lifestyle as well as the availability of passive type leisures provides favourable incentives for such a behaviour.

stance, despite its health benefits, exercising today will always seem less appealing than exercising tomorrow. Procrastination may be one way of looking at the problem nevertheless, another line of research on peer effects showed that individuals habits are also generally influenced by habits of others in the society namely peers. A physically inactive individual may thus encourage others to adopt a similar behaviour. While the lack of willpower might shape an individual's exercising decision, the presence of peer effects amplifies the impact at the global level. The impact of peers on physical habits can flow directly through the impact of peers' habits or indirectly through the impact of peers on time preferences. Thus peers may be producing their impacts by emulation or by reducing (exacerbating) the impact of procrastination on exercising habits or a combination of both.[5] In either contexts, analyzing physical activity habits on an individual basis may not give an accurate picture of the situation and may not provide a sound ground for policy intervention.

Externalities created by social networks have recently gained increased interest and credibility in public and health economics. One of the most well known literature on peer effects and health outcomes focused on the presence of peer effects in obesity (Christakis and Fowler, 2007a; Trogdon et al., 2008; Renna et al., 2008).[6] In response to Christakis and Fowler (2007a), Cohen-Cole and Fletcher (2008a) provide evidence of the presence of peer effects in other health outcomes (e.g., pimples and height) in an attempt to prove the existence of implausible contagions. Another line of research on behavioural patterns found that friends had a significant effect on behaviour related to adolescent smoking (Gaviria and Raphael, 2001; Powell et al., 2005; Fletcher, 2010) , alcohol consumption Gaviria and Raphael (2001), illicit drug use (Clark and Loheac, 2007; Gaviria and Raphael, 2001), risky sex (Jaccard et al., 2005) and fast food consumption (Fortin and Yazbeck, 2011). While the presence of peer effects in health outcomes was used by Cohen-Cole and Fletcher (2008a) as a critique to the work of Christakis and Fowler (2007a), the presence of peer effects in physical fitness (Carrell et al., 2011) as well as in behaviours suggests a new direction to explore: the underlying mechanisms. It is possible that the presence of peers effects in health outcomes is driven by a behavioural change such as smoking, exercising (Christakis and Fowler, 2010) or even nutrition. If this is the case, then the presence of peer effects in lifestyle may be considered as a pathway through which peer effects in health outcomes flow.

Most of the peer effect literature on health outcomes has focused on the presence of peer effects in obesity. In addition, the remaining health related peer effect research focused on health related behaviour. Relating peer effects in health related behaviours and actual health outcomes is a dimension that was somehow under-explored. The objective of this paper falls in this precise stream. Firstly, I tackle physical activity habits from social network standpoint and assess whether there are peer effects in adolescents' exercising habits. The social multiplier effect of physical activity habits

[5]Babcock and Hartman (2010) suggests that peer effects that decrease procrastination do not exist yet peer effects that increase procrastination exists.

[6]The validity of the evidence provided to support the of the existence of peer effects in obesity was heavily questioned by Cohen-Cole and Fletcher (2008a) and Lyons (2011).

is estimated using a *linear in means* equation. It relates a teenager's physical activity habits to his peers' mean physical activity habits (*endogenous effect*), his individual characteristics, as well as his peers' mean characteristics (*contextual effect*). Secondly, I investigate the impact of physical activity on self reported health status. This will be estimated using a panel dynamic health production function. This latter relates current teenager's self-rated health status to his lagged health status and his current physical activities. The peer effect model when combined with the health production function, will allow me to depict the impact of a physical activity's exogenous shock on health outcome through the physical activity's multiplier effect.

The size and the nature of peer effects in exercising habits of adolescents is important both on an individual and public level. First, exercising habits are more likely to be fostered before adulthood (Yang et al., 1999) and affect long term lifestyles as well as health outcomes (Kuh and Ben-Shlomo, 2004). At an early age, parents' education plays an important role. But when a child hits adolescence friends' influence become more important; the desire for conformity effect may often dominate over the parents' education effect. Having a friend who engages in physical activity may be encouraging: it is *better with a buddy* (Jago et al., 2010).[7] Second, there is an important literature that points to the benefits of physical exercising on teenage physical (including obesity) and mental health.[8] All this makes exercising *per se* important as well as desirable, at least at the recommended levels by WHO.[9] Third, understanding the role of peers in adolescent physical activity habits is policy relevant. Preventive public policies affecting youths' health and human capital may be more cost effective and may have a greater impact than alternative redistributive welfare policies targeting adult stages of life. Finally, in the presence of externalities there is room for non-distortionary government intervention. Recent field experiments show that financial incentives interfere in people's decision to exercise (Charness and Gneezy, 2009) and that these incentives may produce spill over effects within a social network (Babcock and Hartman, 2010). This suggests that, in the presence of peer effects, the propagation of the impact of such financial incentives in the society can be expansionary through the effect of the social multiplier.

The estimation of the linear in means model presents well known identification issues that were pointed by Manski (1993) and discussed among others in Moffit (2001), Brock and Durlauf (2001),Durlauf (2004) and Blume et al. (2010). First, it is difficult to separate the overall effect of peers (endogenous + contextual), often referred to as *social effects*, from the *correlated effects i.e.,* effects resulting from the fact that individuals from the same group tend to behave similarly because they are alike (homophily) or share a common environment. Second, it difficult to disentangle *endogenous peer effect* (peer actions) from *contextual peer effects* (peer characteristics). This is referred to as the *Reflection Problem* (see Manski 1993). The identification of endogenous peer effects

[7]Conversely having sedentary friends will encourage him to lead a sedentary lifestyle.

[8]For a complete review refer to section 2.

[9]There is a threshold beyond which exercising might do more harm than good. The optimal frequency depends on the intensity of the activity as well as individual characteristics.

is of crucial importance as it is necessary for the computation of the social multiplier.

To address the first identifications issue, I assume that teenagers' interaction occurs at *network level* and that their physical activity decision is affected by that of their friends in the network. Each school forms a network. I deal with the correlated effects by considering them as fixed at the *network level* and perform a panel-like *within* transformation. The introduction of fixed effects at the school level captures correlated factors associated with network invariant unobserved variables (e.g., similar preferences due to self-selection in schools, same recreational space). In each network, the formation of friendship links is allowed to be stochastic and endogenous. Yet, it becomes strictly exogenous once I condition on school fixed effects, observable individual characteristics and contextual variables. As far as the reflection problem is concerned, disentangling the impact of peer actions from that of peer characteristics is possible (under certain conditions) when the first identification problem is properly handled. Results obtained by Bramoullé, Djebbari and Fortin (2009) show that if linked individuals are separated by a distance 3 within a network (i.e., two strangers linked to each others by two friends), then peer effects are identified. This provides an exclusion restriction: a friends' friends characteristic can be used as an instrument for the friend's physical activity. I estimate the model using a maximum likelihood approach as in Lee et al. (2010) and Lin (2010). I also estimate the model with generalized spatial two stage least square (GS-2SLS) proposed by Kelejian and Prucha (1998) and refined in Lee (2003).

The estimation of the health production function presents also some challenges. An individual's health outcome and physical activity may be endogenous. Following Arellano and Bond (1991), I eliminate individual fixed effects by differencing and use lagged values of physical activity and health outcomes as instruments. I estimate the production function using instrumental variable (IV) approach.

This paper relies on the three waves of the National Longitudinal Study of Adolescent Health (Add Health). I define peers as the group of individuals who are nominated as friends within the same school. Physical activity habits are depicted using three variables: exercising, recreational physical activity and active sports. Each physical activity is reported in terms of frequency (number of times) at which the teenagers has performed each of type of physical activity. The health outcome variable used for the estimation of the health production function is self reported health status. Results from the peer effect model show that there is a positive significant peer effect in certain types of physical activity (exercising) among teenagers in general. Based on the SAR model the estimated social multiplier is 1.12. As for the production function, the estimated parameters show that exercising has a significant effect on improving ones health status of 0.09. These results show that the importance of the impact of physical activity on health outcomes goes beyond its impact on obesity (the most pre-occupying health problem at the present time). It also provides evidence that the presence of peer effects in health outcomes may be mediated by behavioural variables that can be used for

policy intervention. Combined together, these results show that, at the network level, an additional time spent exercising per week increases health status by 5% on average. Thus, physical activity seem to be of a significant importance for health and should not be neglected when considering available preventive measures to decrease healthcare spending.

The remaining of this paper unfolds as follows. Section two provides a review of the related literature. Section three sets the theoretical framework and the empirical methodology. Section four presents the data and some descriptive statistics. Section five discusses the results and section six concludes.

3.2 Review of Literature

The objective of this paper is to assess the presence of peer effects adolescents' physical activity in an attempt to emphasize the presence of a behavioural channel by which peer effects in health outcomes may flow. The existence of such peer effects depends on : (a) the presence of evidence on the existence of peer influence in physical activity, (b) the presence of evidence in support of the impact of physical activity on health outcomes.

Most of the peer effect literature on health outcomes has focused on the presence of peer effects in obesity (Christakis and Fowler, 2007a; Trogdon et al., 2008; Renna et al., 2008). The remaining peer effect literature in the health field focused on health related behaviours (e.g., cigaret consumption). While the literature in social psychology has already pointed at the importance of the role of social networks in individuals' health outcomes (Cohen, 2004), the health economic literature that tackles peer effects in health outcomes and wellbeing remains relatively scarce. A very recent paper by Carrell et al. (2011) uses students from the US military force academy to estimate the impact of peer groups on physical education score.[10] Their sample consists of students that are randomly assigned to the groups. Students are supposed to interact only within these groups during their first year, in addition group members' rooms are in the same dorm and adjacent. In their paper they find evidence that there are peer effects in fitness scores.

As for the literature on peer effect in physical activity, it remains also modest and very recent. An epidemiological study carried by Jago et al. (2010) using data from Bristol 3P project where participants were children in grade 6 recruited from 40 primary school. The physical activity was assessed using accelerometers that the children had to wear for 5 consecutive days. The participating children were asked to identify their best friends as well as the frequency at which they engage in physical activity with their

[10]The fitness core is a measure of physical fitness which is considered a measure of physical health.

nominated (same sex) friend. Their study shows that a child who engages in physical activity with his best friend had higher levels of physical activity. Despite their use of precise measure of physical activity, their analysis has a number of limitations. First, it was conducted separately for boys and girls leaving out across-sex interactions.[11] It also ignores the simultaneity problem accruing that the physical activity intensity between two friends can be due to the fact that they are exposed to the same environment. In addition, the simultaneity problem between the outcome of both friends is not addressed. Finally, by focusing on best friends, they narrow down the definition of social networks such that it does not reflect well the reality of social interactions.

In the health economics literature, Babcock and Hartman (2010) used a partial population approach to investigate the size of a financial incentives' impact on exercising behaviour through its impact on peers. In their paper, they randomly selected college students who shared a common residence hall and collected friendship data for students within the experiment. These students were then randomly attributed to control and treatment groups. Those who were treated were paid the amount of 80$ for visiting the gym at least 8 times every four weeks.[12] The obtained results reveal that treated students are affected by their treated friends and that this is effect is increasing in the number of friends who are treated. Also, it appears that there is no impact for untreated students regardless of the treatment status of their friends. This suggests the absence of peer effects in the absence of incentives. This analysis raises some issues. First, the exercising frequency is assessed by card swiping and the payment was conditional on having the card swiped. Thus observing peer effects in the treated group only could be attributed to the fact that treated friends were only swiping their cards together. In addition, the control group was very aware of the treatment. This might affect the behaviour of the control group which could have been different without this information.

Another study by Yakusheva et al. (2010) used an experimental approach to assess peer effects in weight gain and weight management behaviour. In their natural experiment they used freshman students of a private Midwestern university were students were randomly assigned. They examined whether students' weight is affected by that of their assigned roommate. The authors also investigate weight gain behavioural variables (i.e., eating habits, exercising, use of weight loss supplements). They find evidence in favour of the presence of significant peer effects in exercising outdoors. The endogenous effects were estimated to be 0.13. This means if a freshman student's roommate increases his exercising outside by one time this will lead to an increase in his roommate's exercising by 0.13 times per week.

The above mentioned experiments raise a number of limitations. First, they focused on a limited sample (*e.g.*, best friends, children ages 10 to 11 years, freshman students). Furthermore, their estimates are likely to underestimate social interaction effects as best friends or roommates do not represent the *true* social network shaping the adolescent's

[11]The data on cross-sex was very scarce this is why the authors chose to focus on same sex friends.
[12]The visits were calculated as the number of times the card was swiped in the gym.

behaviour (Carrell et al., 2008; Stinebrickner and Stinebrickner, 2006a). This paper, unlike most of the papers on peer effects and physical activity, relies on observational non-experimental data where social networks reflects, to a reasonable extent, the true social network of an adolescent. An important part of adolescents' social interactions occur in schools, thus, by considering school friends I can capture their real social network quite reasonably. The use of non-experimental data allows me exploit the rich social networks available, through the use of a structural approach (linear-in-means), when estimating the impact of peer effects. This, in turn, allows me to exploit the similarities between the linear-in-means model and the spatial autoregressive approach developed in spatial econometrics.

The recent rising interest in the presence of peer effects in physical activity is motivated by the public concern with the obesity epidemics. The relative importance of the role of physical activity in prevention and treatment of obesity is still on the debate table. Part of the literature finds that physical activity plays an important role in the obesity epidemics, while another part of the literature points that eating habits play a central role in the recently observed weight gains trends.[13] Yet, the importance of the impact of physical activity on health outcomes goes beyond its impact on obesity. Physical activity has a positive impact on physical and mental health outcomes.

There is a great body of health literature on the benefits of physical activity in preventing chronic diseases such as coronary and heart disease, colorectal cancer and diabetes. The inverse relationship between physical activity and heart attack was pointed by Paffenbarger et al. (1978). According to Magnus et al. (1979) regular daily activity such as gardening, walking or cycling were associated with a lower risk of acute coronary disease. The preventive impact of physical activity appears to be independent from family history, the level of health concern, smoking, weight and physical stature (Morris et al., 1990). Also, a series of meta-analysis on the preventive role of physical activity on coronary heart disease show that physical activity reduces the risk of coronary and heart incident as well as cardiovascular and all-cause mortality (e.g., Powell et al., 1987; Berlin and Colditz, 1990; Sofi et al., 2008; Nocon et al., 2008). More specific studies such as Hamer and Chida (2008) found that walking plays an important role in reducing the risk of coronary heart disease. Zheng et al. (2009) find evidence in the same direction and point that physical activity should be considered as a prevention measure.

Besides its impact on one of the most costly health problem, physical activity seem to reduce the risk of recurrence and mortality for stage III colorectal cancer patients (Meyerhardt et al., 2006a,b; Haydon et al., 2006). These results are robust to any type of chemo-therapy received, gender and age of patients as well as their BMI.[14] In addition, physical activity appears to help preventing colon cancer (for a complete

[13]Anderson et al. (2003), Andersen et al. (1998), Ruhm (2000), Courtemanche (2007), Bleich et al. (2007), Fortin and Yazbeck (2011), Lakdawalla and Philipson (2002).

[14]Note that Haydon et al. (2006) has found that BMI has an effect.

review see Samad et al., 2005; Wolin et al., 2009).

Physical inactivity has an important role in increasing the risk of type 2 diabetes. There is a large body of literature on the impact of physical activity on diabetes prevention (Li et al., 2006b) and risk. Helmrich et al. (1991) highlight the presence of an inverse relationship between physical activity (e.g., walking) and type 2 diabetes. Other studies using data from Nurses Health study (Hu et al., 1999, 2001) and data from Iowa Women's Health Study (Folsom et al., 2000) show similar results. The impact of physical activity on reducing the risk of type two diabetes has been confirmed outside the U.S too. Hu et al. (2003) and Hu et al. (2004) find similar association using Finnish data. Also, Okada et al. (2000) using Osaka Health Survey (Japan) finds that moderate exercise on regular basis and a once a week vigorous exercise decrease the risk of type 2 diabetes.

Besides its impact on diagnosed disease, physical activity has an impact on health status. The impact of physical activity on health has also been tackled from a general perspective using self reported health status (physical and mental) and physical activity. Papers by Warburton et al. (2006) and Pisinger et al. (2009) show that physical activity has a positive impact on health status. My paper finds its basis in this literature, as well as the literature on peer effects and physical activity. It assesses whether it is possible to depict peer effects in health outcomes. It specifically focuses on self-reported health status. While my paper is closely related to the work of Babcock and Hartman (2010) and Yakusheva et al. (2010) in its interest in the presence of peer effects in physical activity, it remains quite distinct in many respects as explained earlier in this section. An additional originality of this paper lies in the fact that it relates physical activity to health status in an attempt to highlight a potential mechanism through which peer effects in health outcomes are mediated.

3.3 Structural Model and Estimation Methods

This section has two main parts. First, I present a linear-in-means based model to estimate the presence of peer effects in physical activity.[15] I linearly relate the outcome of an individual to his own characteristics, the corresponding mean characteristics of his peers as well as their mean outcome. Unlike the linear-in-means model I define social interactions at the network level. This, in turn, allows the identification of endogenous peer effects (physical activity of peers) and exogenous peer effects (characteristics of peers) under specific conditions. I also present the different estimation methods used to

[15]This structural model can be derived from a choice-theoretic approach where each individuals physical activity is obtained from the maximization of his quadratic utility function which depends on his individual characteristics, his physical activity and his reference group's mean physical activity and mean characteristics. This approach also assumes that social interactions have reached a noncooperative (Nash) equilibrium at which expected physical activity are realized.

estimate this model. Second, I present the dynamic health production function which relates an adolescent health status to his physical activity. I also present the estimation method that I follow to estimate this model.

3.3.1 A structural model of peer effects in physical activity

Suppose that I have a set of N individuals i that are partitioned in a set of L networks. A network is defined as a structure (e.g., school) in which individuals are tied by a certain affinity (e.g., friendship). Each individual i in the network has a set of nominated friends N_i of size n_i that constitute his reference group (or peers). I assume that i is excluded from his reference group. Since peers are defined as nominated friends, the number of peers will not be the same for every network member. Let $\mathbf{G}_l$ $(l = 1, \ldots, L)$ be the social interaction matrix for a network l, its element g_{lij} takes a value of $\frac{1}{n_i}$ for i's nominated friend and zero otherwise.[16] Further, I assume that each individual interacts only with his friends in the same network and with no other friends outside the network. I define y_{il} as the physical activity undertaken by individual i in network l, x_{il} represents the individual i's observable characteristics, $\mathbf{y}_l$ the vector of physical activity in network l, and $\mathbf{x}_l$ is the corresponding vector for individual characteristics. To simplify my presentation, I look at only one characteristic (*e.g.*, adolescent's age).[17] The correlated effects are captured through α_l. The ε_{il}'s are the error terms, they capture i's unobservable characteristics that are not fixed within the network. Formally, one can write the linear-in-means model for individual i as follows:

$$y_{il} = \alpha_l + \beta \frac{\sum_{j \in N_i} y_{lj}}{n_i} + \gamma x_{li} + \delta \frac{\sum_{j \in N_i} x_{lj}}{n_i} + \varepsilon_{li}, \tag{3.1}$$

where $\frac{\sum_{j \in N_i} y_{lj}}{n_i}$ and $\frac{\sum_{j \in N_i} x_{lj}}{n_i}$ are respectively his peers' mean physical activity frequency and characteristics. In the context of my paper, β is the *endogenous social effect.* It reflects how the individual's frequency of physical activity is affected by his peers' mean frequency of physical activity. One usually assumes that $|\beta| < 1$. The *contextual social effects* are represented by the parameter δ. It captures the impact of his peers' mean characteristic on his physical activity. It is important to note that the matrices of $\mathbf{G}_l$'s and the vectors of $\mathbf{x}_l$'s are stochastic but strictly exogenous conditional on α_l , that is, $E(\varepsilon_{li}|\mathbf{x}_l, \mathbf{G}_l, \alpha_l) = 0$. This assumption is flexible enough to allow for correlation between the network's unobserved common characteristics (e.g., school's recreational facilities quality) and observed characteristics (e.g., parent's education).[18] Nevertheless, once I condition on these common characteristics, parent's education is assumed to be independent of i's idiosyncratic unobserved characteristics. Following the formulation of Bramoullé et al. (2009), let $\mathbf{I}_l$ be the identity matrix for a network l and $\boldsymbol{\iota}_l$ the

[16] The $\mathbf{G}_l$ matrix is row normalized.

[17] The model can be easily generalized using more than one characteristic K >1.

[18] In this case $\mathbb{E}(\alpha_l|\mathbf{G}_l, \mathbf{x}_l) \neq 0$.

corresponding vector of ones, I can rewrite (3.1) in matrix notation as follows:

$$\mathbf{y}_l = \alpha_l \boldsymbol{\iota}_l + \beta \mathbf{G}_l \mathbf{y}_l + \gamma \mathbf{x}_l + \delta \mathbf{G}_l \mathbf{x}_l + \boldsymbol{\varepsilon}_l. \tag{3.2}$$

Since $(\mathbf{I}_l - \beta \mathbf{G}_l)$ is invertible then, in matrix notation, the reduced form of the model for the network l can be written as follows:

$$\mathbf{y}_l = \alpha_l (\mathbf{I}_l - \beta \mathbf{G}_l)^{-1} \boldsymbol{\iota}_l + (\mathbf{I} - \beta \mathbf{G}_l)^{-1} (\gamma \mathbf{I}_l + \delta \mathbf{G}_l) \mathbf{x}_l + (\mathbf{I}_l - \beta \mathbf{G}_l)^{-1} \boldsymbol{\varepsilon}_l, \tag{3.3}$$

To address problems that arise from the fact adolescents are sharing the same environment, I perform a panel-like *within* transformation to the model.[19] More precisely, I average equation (3.3) over all students of network l and subtract it from i's equation.[20] Let $\mathbf{I}_l - \mathbf{H}_l$ be the matrix that obtains the deviation from network means such that $\mathbf{H}_l = \frac{1}{n_l}(\iota_l \iota_l')$. The network within transformation will eliminate the correlated effects α_l over all the networks of the sample. Let $\mathbf{K}$ be a block-diagonal matrix with $\mathbf{K}_l = \mathbf{I}_l - \mathbf{H}_l$ on its diagonal and $\mathbf{G}$ be a block-diagonal matrix with $\mathbf{G}_l$ on its diagonal, then writing the model in a more compact form defined over all networks yields the following equation:

$$\mathbf{K}\mathbf{y} = \mathbf{K}(\mathbf{I} - \beta \mathbf{G})^{-1} (\gamma \mathbf{I} + \delta \mathbf{G}) \mathbf{x} + \mathbf{K}(\mathbf{I} - \beta \mathbf{G})^{-1} \boldsymbol{\varepsilon}. \tag{3.4}$$

Identification

The structural model that I am using presents two main identification issues:

a-Simultaneity

The reflexion problem pointed by Manski (1993) may hinder the separation of the contextual effects from the endogenous effects. This identification issue was analyzed by Bramoullé et al. (2009) who tackled this issue in the context of network based interactions. In the absence of correlated effects (i.e., $\alpha_l = \alpha$, $\forall\ l$), Bramoullé et al. (2009) show that it is possible to identify the structural parameters of the peer effect model if the matrices $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ are linearly independent (link distance =2). This implies the existence of at least two adolescents who are strangers to each other, but who are linked by a friend (intransitive triads). The identification condition becomes more

[19] Alternatively, one could perform a *local* transformation where the averaging is performed only over all student i's friends.
Bramoullé et al. (2009) has shown that if the model is identified with a global transformation, it will also be identified with a local transformation.

[20] This transformation (i.e., *gobal* transformation) imposes less restrictive identification conditions. Alternatively, one could perform a *local* transformation where the averaging is performed only over all student i's friends.
Bramoullé et al. (2009) has shown that if the model is identified with a global transformation, it will also be identified with a local transformation.

restrictive in the presence of fixed network effects. According Bramoullé et al. (2009), a sufficient condition for the identification of the structural parameters is the linear independence between $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ and $\mathbf{G}^3$ (link distance=3). This condition is satisfied when two adolescents who are strangers to each other are linked by two friends.

b-Endogenous Network formation

Another identification problem may arise due to self-selection into school networks. This might introduce serious bias in estimating overall peer effects (endogenous+ exogenous) if the source driving this selection process are not fully observable. More specifically, the correlations between the unobserved network specific factors and the regressors will be the principle source driving this bias. In this paper, it is assumed that network-specific fixed effects do capture all these factors. This is consistent with a two step link formation procedure. In the first step, each adolescent self select into a school. This selection bias will be due to the school specific characteristics. In the second step, the adolescent forms links in the school. These links can be formed either randomly or based on observable characteristics only. As mentioned earlier the network fixed effects are dealt with through a panel within transformation.

Other types of correlated effects can occur for reasons other than common environmental factors. For instance, one can think of other uncommon unobserved reasons for which some people might group together. To account for this possibility, I allow for spatial autocorrelation in the error term in (3.3) such that:

$$\boldsymbol{\varepsilon}_l = \rho \mathbf{G}_l \boldsymbol{\varepsilon}_l + \boldsymbol{\xi}_l, \tag{3.5}$$

where the innovations, $\boldsymbol{\xi}_l$ are assumed to be i.i.d. $N(0, \sigma^2 \mathbf{I}_l)$ and $0 < \sigma^2 < b$ with b $< \infty$. The autoregressive spatial parameter ρ is assumed to be less than one and the matrix $(\mathbf{I} - \rho \mathbf{G}_l)$ is assumed to be nonsingular $\forall\ |\rho| < 1$, $l = 1, ..., L$ (Kelejian and Prucha, 1999). Given these assumption I can write:

$$\boldsymbol{\varepsilon}_l = (\mathbf{I}_l - \rho \mathbf{G}_l)^{-1} \boldsymbol{\xi}_l. \tag{3.6}$$

For a given vector of outcomes $\mathbf{y}$ defined for all school's networks and a corresponding matrix of individuals' characteristics $\mathbf{X}$, the model is given by the following compact structural form:

$$\mathbf{KMy} = \beta \mathbf{KMGy} + \mathbf{KMX}\boldsymbol{\gamma} + \mathbf{KMGX}\boldsymbol{\delta} + \boldsymbol{\nu}, \tag{3.7}$$

where $\mathbf{M} = (\mathbf{I} - \rho \mathbf{G})$ and $\boldsymbol{\nu} = \mathbf{K}\boldsymbol{\xi}$. As previously highlighted, in the presence of correlated effects, the parameters β, γ and δ are identified if the matrices $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ and $\mathbf{G}^3$ are linearly independent, a condition that will be checked with the data I use.

The elimination of fixed effects using a *within* transformation leads to singular variance-covariance matrix such that $E(\boldsymbol{\nu}\boldsymbol{\nu}' \mid \mathbf{X}, \mathbf{G}) = \mathbf{K}\mathbf{K}'\sigma^2 = \mathbf{K}\sigma^2$. To resolve

this problem I compute the orthonormal base of $\mathbf{K}_l$. Let $\mathbf{Q}'_l$ the sub-matrix corresponding to the eigenvalues of 1. $\mathbf{Q}'_l$ has the following properties : $\mathbf{Q}'_l\mathbf{Q}_l = \mathbf{I}_{n_l^*}$, $\mathbf{Q}_l\mathbf{Q}'_l = \mathbf{K}_l$ and $\mathbf{Q}'_l\iota = 0$ where $n_l^* = n_l - 1$. Let us denote by $\mathbf{Q}$ the block-diagonal matrix, where $\mathbf{Q}'_l$ are the matrices on the diagonal.Given that $\mathbf{G}$ is row-normalized I also have $\mathbf{Q}'\mathbf{M} = \mathbf{Q}'\mathbf{M}\mathbf{Q}\mathbf{Q}' = \mathbf{M}^*\mathbf{Q}'$ with $\mathbf{K}\mathbf{M} = \mathbf{Q}\mathbf{Q}'\mathbf{M}$ this means that $\mathbf{K}\mathbf{M} = \mathbf{Q}\mathbf{M}^*\mathbf{Q}'$ and that $\mathbf{K}\mathbf{M}\mathbf{G} = \mathbf{Q}\mathbf{M}^*\mathbf{Q}'\mathbf{G} = \mathbf{Q}\mathbf{M}^*\mathbf{G}^*\mathbf{Q}'$. Pre-multiplying (3.7) by $\mathbf{Q}'$, the model can now be written as follows:

$$\mathbf{M}^*\mathbf{y}^* = \beta\mathbf{M}^*\mathbf{G}^*\mathbf{y}^* + \mathbf{M}^*\mathbf{X}^*\gamma + \mathbf{M}^*\mathbf{G}^*\mathbf{X}^*\delta + \boldsymbol{\nu}^*, \tag{3.8}$$

where $\mathbf{y}^* = \mathbf{Q}'\mathbf{y}$, $\mathbf{X}^* = \mathbf{Q}'\mathbf{X}$, $\mathbf{G}^* = \mathbf{Q}'\mathbf{G}\mathbf{Q}$ and $\boldsymbol{\nu}^* = \mathbf{Q}'\xi$. Note that with this transformation one has $E(\boldsymbol{\nu}^*\boldsymbol{\nu}^{*\prime} \mid \mathbf{X}, \mathbf{G}) = \sigma^2\mathbf{I}_{n_l^*}$. Note also that while the block-diagonal matrix $\mathbf{G}$ is composed of row normalized matrices that have zero diagonals ($\mathbf{G}_l$), $\mathbf{G}^*$ will not preserve these properties. This come without a cost as far as the consistency of the estimates is concerned.

Following Lee et al. (2010), I propose two approaches to estimate the peer effect model presented earlier: (1) a maximum likelihood approach (ML), and (2) a generalized spatial two stage least squares approach (GS-2SLS).

Maximum Likelihood (ML)

Using the same notation as in the previous section, and assuming that the errors are normally distributed the likelihood function can be expressed as follows[21]:

$$\ln \mathbb{L} = \frac{-n^*}{2}\ln(2\pi\sigma^2) + \sum_{l=1}^{L}\ln|\mathbf{I}_{n_l^*} - \beta\mathbf{G}_l^*| + \sum_{l=1}^{L}\ln|\mathbf{I}_{n_l^*} - \rho\mathbf{M}_l^*| - \frac{1}{2\sigma^2}\sum_{l=1}^{L}\boldsymbol{\nu}_l^{*\prime}\boldsymbol{\nu}_l^*, \tag{3.9}$$

where $n^* = \sum_{l=1}^{L} n_l^* = N - L$, and, from (3.8), $\boldsymbol{\nu}_l^* = \mathbf{M}_l^*(\mathbf{y}_l^* - \beta\mathbf{G}_l^*\mathbf{y}_l^* - \mathbf{X}_l^*\gamma - \mathbf{G}_l^*\mathbf{X}_l^*\delta)$. Maximizing (3.9) with respect to $(\beta, \boldsymbol{\gamma}', \boldsymbol{\delta}', \rho, \sigma)$ yields the maximum likelihood estimators of the model.[22] Interestingly, the ML method is implemented after the elimination of the network fixed effects. Therefore, the estimators are not subject to the incidental parameters problem that may arise since the number of fixed effects increases with the the size of the networks sample.

[21] It is important to note that we do not account for the count nature of the dependant variable. We expect that this is bias our estimates downwards.

[22] For computational simplicity, one can concentrate the log-likelihood function (3.9) and maximize the concentrated log-likelihood function. See Lee et al. (2010) for more details.

Generalized spatial two stage least squares (GS-2SLS)

To estimate the model in (3.8) I adopt a generalized spatial two-stage least squares procedure presented in Lee et al. (2010). This approach provides a simple and tractable numerical method to estimate an asymptotically optimal IV estimator. In the case of my paper this method will consist of a two-step estimation.[23] To simply the notation, let us denote by $\tilde{\mathbf{X}}^*$ the matrix of explanatory variables such that $\tilde{\mathbf{X}}^* = [\mathbf{G}^*\mathbf{y}^* \quad \mathbf{X}^* \quad \mathbf{G}^*\mathbf{X}^*]$. Let $\mathbf{P}$ be the lighting matrix such that $\mathbf{P} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$, and $\mathbf{S}$ a matrix of instruments such that $\mathbf{S} = [\mathbf{X}^* \quad \mathbf{G}^*\mathbf{X}^* \quad \mathbf{G}^{*2}\mathbf{X}^*]$. In the first step, I estimate the following 2SLS estimator:

$$\hat{\boldsymbol{\theta}}_1 = (\tilde{\mathbf{X}}^{*\prime}\mathbf{P}\tilde{\mathbf{X}}^*)^{-1}\tilde{\mathbf{X}}^{*\prime}\mathbf{P}\mathbf{y}^*,$$

where $\hat{\boldsymbol{\theta}}_\mathbf{1}$ is the first-step 2SLS vector of parameters $[\hat{\gamma}' \quad \hat{\delta}' \quad \hat{\beta}]$ of the structural model.

In the second step I estimate a 2SLS using $\hat{\mathbf{Z}}$ as instruments. To avoid any confusion with the first step, let us denote this new matrix of instruments $\hat{\mathbf{Z}}$ such that $\hat{\mathbf{Z}} = \mathbf{Z}(\hat{\boldsymbol{\theta}}_1)$ with:

$$\hat{\mathbf{Z}} = [\mathbf{G}^*\hat{\mathbf{y}}^* \quad \mathbf{X}^* \quad \mathbf{G}^*\mathbf{X}^*],$$

and where

$$\mathbf{G}^*\hat{\mathbf{y}}^* = \mathbf{G}^*(\mathbf{I}-\beta\mathbf{G}^*)^{-1}[(\mathbf{X}^*\gamma + \mathbf{G}^*\mathbf{X}^*\delta)].$$

I then estimate:

$$\boldsymbol{\theta}^{LEE} = (\hat{\mathbf{Z}}'\tilde{\mathbf{X}}^*)^{-1}\hat{\mathbf{Z}}\mathbf{y}^*.$$

This estimator can be shown to be asymptotically best IV estimator. We also have:

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{Lee}-\boldsymbol{\theta}) \sim N\left(0, \left[\mathbf{Z}'\tilde{\mathbf{X}}^*\mathbf{R}^{-1}\tilde{\mathbf{X}}^*\mathbf{Z}\right]^{-1}\right).$$

The matrix $\mathbf{R}$ is consistently estimated by

$$\hat{\mathbf{R}} = s^2\frac{\hat{\mathbf{Z}}'\hat{\mathbf{Z}}}{N},$$

where $s^2 = N^{-1}\sum_{i=1}^{N}\hat{u}_i{}^2$ and $\hat{u}_i$ are the residuals from the second step. It is important to note that, as in Kelejian and Prucha (1998), I assume that errors are homoscedastic. The estimation theory developed by Kelejian and Prucha (1998) under the assumption of homoscedastic errors does not apply if I assume heteroscedastic errors (Kelejian and Prucha, 2010).

[23] Note that for this particular case I impose $\rho = 0$ and thus $M_l = I_l$

3.3.2 A health production function

In this section I augment the physical activity peer effect model with a health production function by exploiting the panel data available on physical activity and health outcomes. Specifically, I relate an individual's self-reported health outcomes in time t to his lagged health outcome, his physical activity as well as his own characteristics in period t. Let y_{it}^h be an individual's health outcome at time t, and y_{it}^a be the individual's physical activity. Thus, for a given vector of characteristics $\mathbf{x}_{it}$, the health production function can written as follows (for notational simplicity I suppress l):

$$y_{it}^h = \pi_0 + \pi_1 y_{i,t-1}^h + \pi_2 y_{it}^a + \boldsymbol{\pi}_3 \mathbf{x}_{it} + \varepsilon_{it} \tag{3.10}$$

where, The error term ε_{it} can be decomposed into a time invariant term μ_i and a serially uncorrelated idiosyncratic term ν_{it}. The time invariant component of the error term will be considered as a fixed effect. The reduced form obtained from (3.1) and (3.10) in stacked form allows me to depict the magnitude of a change in health resulting from an exogenous shock on physical activity via the social multiplier. Thus, plugging in the peer effect model in the health production function yields the following equation:

$$E[\mathbf{y}_t^h|\mathbf{y}_{t-1}^h,.] = \pi_0 + \pi_1 \mathbf{y}_{t-1}^h + \pi_2 \left[\frac{\alpha}{1-\beta} + \frac{\gamma+\delta}{1-\beta}\bar{\mathbf{x}}_t\right] + \boldsymbol{\pi}_3 \mathbf{x}_t + E[\varepsilon_{it}|\mathbf{y}_{t-1}^h,.] \tag{3.11}$$

Partially differentiating equation (3.11) with respect to α will yield $\frac{\pi_2}{1-\beta}$ which reflects the short-run impact in health satisfaction units of an exogenous shock on physical activity. The resulting short run parameter $(\frac{\pi_2}{1-\beta})$ entails two components: the impact of the physical activity on the health (i.e., π_2) and the multiplier effect $(\frac{1}{1-\beta})$. In the long run $\mathbf{y}_t^h = \mathbf{y}_{t-1}^h$ thus, the long-run impact in health units of an exogenous shock on physical activity will be: $\frac{\pi_2}{1-\beta}\frac{1}{1-\pi_1}$.

Estimating (3.10) with a simple OLS will not be consistent for two reasons. First, reverse causality might occur; an individual's physical activity is not exogenously determined and may be affected by his own health status. In addition, there is a correlation between lagged dependant variable and the error term due the presence of a time invariant error component. One way to resolve this issue is to apply a first difference. While such a transformation wipes out all individual fixed effects (except age) it has the advantage of making the correlation between right hand side regressors and the error term easier to handle. Formally the transformed model can be written as follows:

$$\Delta y_{it}^h = \pi_0 + \pi_1 \Delta y_{i,t-1}^h + \pi_2 \Delta y_{it}^a + \Delta \nu_{it} \tag{3.12}$$

To resolve the correlation between the right hand side variables and the error term one can instrument for $\Delta y_{i,t-1}^h$ and Δy_{it}^a. To instrument the lagged dependant variable, I can either use $\Delta y_{i,t-2}^h$ or $y_{i,t-2}^h$ (Hsiao ,1981). Following the recommendations of Arellano and Bond (1991), I instrument using y_{t-2}^h. As for the physical activity variable y_{it}^a, I

instrument using $\Delta y^a_{i,t-2}$ as well as the strictly exogenous variables that I have used in the previous section.

3.4 Data and Descriptive Statistics

3.4.1 Data

The Add Health survey is a longitudinal study that is nationally representative of American adolescents in grades 7 through 12. It is one of the most comprehensive health surveys that contains fairly exhaustive social, economic, psychological and physical well-being variables along with contextual data on the family, neighbourhood, community, school, friendships, peer groups, romantic relationships, *etc.* In the first wave (September 1994 to April 1995), all students (around 90 000) attending the randomly selected high schools were asked to answer a short questionnaire. An in-home sample (core sample) of approximately 20 000 students was then randomly drawn from each school. These adolescents were asked to participate in a more extensive questionnaire where detailed questions were asked. Information on (but not limited to) health, nutrition, expectations, parents' health, parent-adolescent relationship and friends nomination was gathered.[24] This cohort was then followed in-home in the subsequent waves in 1996 (wave II) and 2001 (wave III). The extensive questionnaire was also used to construct the saturation sample that focuses on 16 selected schools (about 3000 students). Every student attending these selected schools answered the detailed questionnaire. There are two large schools and 14 other small schools. All schools are racially mixed and are located in major metropolitan areas except one large school that has a high concentration of white adolescents and is located in a rural area.

In this paper I use the saturation sample of wave II in-home survey to investigate the presence of peer effects in physical activity.[25] This allows me to depict physical activity patterns of each adolescent and relate it to that of his peer group. In addition, the availability of friend nomination allows me to retrace school friends and thus construct friendship networks.[26] To estimate the health production function, I considered information from wave I, wave II and wave III.

I exploit friends nominations to construct the network of friends. Thus, I consider all nominated friends as network members regardless of the reciprocity of the nomination. If an adolescent nominates a friend then a link is assigned between these two adolescents but the link does not need to be reciprocal.

[24] Adolescents were asked to nominate either 1 female friend and 1 male friend or 5 female friends and 5 male friends.

[25] It includes three types of physical activity exercising, active sports and recreational activities.

[26] In the saturation sample, all students are asked to name their friends.

3.4.2 Dependant variables

In the estimation of the peer effects model, I use three key variables: exercising, playing active sports and recreation sports. Exercising refers to activities that are mainly cardio-vascular such as jumping rope or jogging. Playing active sports includes soccer, basketball or swimming and this is basically team sports. As for recreational sports, it usually includes all kinds of skating, biking and skying. Ideally, one should have intensity and frequency. In my case frequency is reported directly in the survey and the information on the intensity of the physical activity is given by the different types of physical activities.

To assess the impact of physical activity on health (i.e., production function), I use self reported health status. The use of self-reported health status has revealed to be reliable in the prediction of mortality as well as functional decline. It is argued that it is a better predictor than diagnosed health problems, biological or life-style risk factors (Idler and Kasl, 1991; McGee et al., 1999; Lee, 2000). Further, Benjamins et al. (2004) and Pisinger et al. (2009) show that self reported health status is weakly associated with accidental death (e.g., suicide, homicide...) and strongly associated with deaths resulting from lifestyle-related disease (e.g., coronary heart disease, cancer, diabetes...).

3.4.3 Descriptive Statistics

The dependant variable of the peer effects model is physical activity. It is reported by three different variables namely exercising, active sports and recreational sports. These three activities differ in terms of their intensity. Their value varies between 0 and 3. For a frequency between one and twice a week the variable takes a value of 1, for a frequency between three to four times the variable takes a value of 2, finally for a frequency is between five and six times a week the variable takes a value of 3.[27]

The mean and standard deviation of the endogenous variable and it covariates are reported in table 3.1. It appears from my data that exercising is the most popular form of physical activity. Only 14% of the adolescents did not exercise at all while 27 % did not play active sports and 62 % did not engage in any recreational sports. Also, 34% of adolescents exercise between once to twice per week while 24% engage in recreational activities.[28]

The covariates of the physical activity equation include the adolescent's personal characteristics, family characteristics as well as the corresponding contextual social effects. The personal characteristics are gender, age, ethnicity (white or other) and

[27]This a common way of reporting physical activity data.

[28]The figures are not reported but can be provided upon request.

grade. I observe that 50% of the sample are females, that the mean age is 16.3 years and that 57% are white. Family characteristics are dummies for mother's presence as well as mother and father education. I observe that 45% of mothers and 44% fathers have at least some college education. To control further for parents' income I use child allowance as a proxy. An adolescent's allowance is on average 8.41 $ per week, around 53% of the adolescents in the sample have a weekly allowance.

As for the health production functions, the dependant variables I use is the self-reported health status. The covariates are the exercise, active sports and recreational activities, the lagged health status variable (Table 3.2), age, and all time-invariant variables appearing in Table 3.2.

3.5 Results

3.5.1 Baseline: OLS peer effects estimates

Using a simple OLS regression, I estimate the physical activity peer effect model. For each type of physical activity, I regress the physical activity of an adolescent on the average physical activity of his group of reference, his individual characteristics as well as the average characteristics of his reference group. To account for correlated effects mentioned earlier, I perform a panel-like within transformation. Note that the naive OLS estimates and the OLS_w estimates are inconsistent. Both of them do not account for the simultaneity problem discussed earlier. Further, the naive OLS ignores the presence of correlated effects resulting from the fact that adolescents in the same reference group tend to chose a similar frequency of physical activity.

Estimation results of the naive OLS and the OLS-within reported respectively in table 3.3 and table 3.4. It appears that there is a significant positive peer effect in physical activity at the general level. The magnitude of these effects vary depending on the type of the activity studied. An adolescent would increase his active sports frequency by 0.24 in response to a 1 unit increase in his friends active sports practice. This corresponds to a 40 % (=0.24/0.59) increase in active sports. This impact becomes a little lower when I apply the within transformation (0.21 or 36%). The recreational sports parameter reveals the presence of peer effects that are lower than those of active sports 0.19 compared to 0.24 for active sports. On average the increase in recreational sports frequency corresponds to an increase that ranges between 13% and 11% when I consider the within estimates. As for the exercise peer effect parameter, it has the lowest magnitude (around 0.17, or 10%) this parameter decreases to 0.12 (or 7 %) when I consider the OLS_w estimates.[29]

[29]How do these results compare to the ones obtained in the literature? To my knowledge, very

Individual characteristics reported in table 3.3 and table 3.4 exhibit fairly the same pattern. Physical activity seem to be decreasing in age for all three type of physical activity and increasing for adolescent who are in grade 9-10 except for recreational sports where the opposite occurs. The negative association between age and physical activity is well established one in the literature the steepest drop has been documented to be between the ages of 13 and 18 (Sallis, 2000). Also, girls are less likely to engage in recreational sports and active sports. Father's education has a positive on active sports participation and being white has a positive effect on exercising and recreational sports. As for contextual peer effects, it seems that friends' age has a negative influence on the mean frequency of active sports and recreational sports. Also the mean active sports increase with the mean education of friends' fathers and mean exercising decreases with mean education of friends' fathers and increases with mean education of friends' mothers.

3.5.2 ML peer effect estimates

Table 3.5 shows maximum likelihood estimates of the linear in means model with school fixed effect where the errors are allowed to be autocorrelated within the networks (GSARAR model). Table 3.6 reports the results of this same model where the error term is assumed to be normally distributed (GSAR). By imposing normality on the error term, rather than autocorrelation, I can move from the first equation to the second. Results reported in table 3.5 and table 3.6 show that the estimated peer effects are between -0.10 and 0.17 (or - 7% and 29%). The estimates are smaller than those obtained using the naive OLS as well as the OLS-within (between 7% and 40%). This difference between the ML estimates and the ones of the previous section are likely to reflect a positive correlation between the error term and physical activity.

Based on the GSARAR model (table 3.5), I notice that the estimated endogenous peer effect parameter is statistically insignificant for all three activities while the autocorrelation parameter is positive and significant for active sports and recreational sports only. The presence of positive significant autocorrelation suggests that adolescents' friendship formation may be due to their common preference for certain types of physical activity such as active sports (e.g., team sports) and recreational sports (e.g., biking). As for exercising, the hypothesis that the error term is correlated is rejected. This, in turn, suggest that the assumption of autocorrelated errors might not be suitable for this type of physical activity. While the absence of endogenous peer effects for active sports and recreational sports suggest that friends might not play an important role in shaping an individual's choice of this type of physical activity, it does not suggest the same for exercising. This is why I turn to a more restrictive model,

little evidence based research has been done on the presence of peer effects in physical activity. All the studies that I am aware of used an experimental design approach to estimate these effects and cannot be used as a reference at this point.

where the autocorrelation parameter is assumed to be null.

Estimation results for the GSAR model (table 3.6) show that all the peer effects parameters are positive and statistically significant (between 0.10 and 0.17). This is in line with the estimates obtained in the literature on peer effect and physical activity. Previous research on peer effect and physical activity has mainly focused on experimental approaches. Whilst the parameters estimated in my model are not exactly comparable to those obtained in the literature, one can still use these latter for guidance. The gym visits peer effect parameter estimated by Babcock and Hartman (2010) is 0.10 and is statistically significant. Also, Yakusheva et al. (2010) finds evidence that allow for the conclusion that having a roommate who exercise outside the gym will increase one own exercising by 0.13 times per week (i.e., on average 11%).

Combining the information from ML estimates in tables 3.5 and 3.6 reveal the presence of significant positive peer effects in exercising of 0.11 (on average 7 %) and non-significant peer effects for active sports and recreational sports. An adolescent will increase his exercise by 0.11 times in response to a one time increase of his friends exercising per week. The associated social multiplier is 1.12 ($= \frac{1}{1-0.11}$). The magnitude of the multiplier does not come as a surprise as the existence of peer effects in physical activity is partially hindered by the presence of salient cost: it is always more appealing to exercise tomorrow instead of today.[30]

3.5.3 GS-2SLS peer effect estimates

Estimation results in table 3.7 are in contrast with the estimates previously obtained. They are statistically insignificant and smaller (0.03 and -0.01) expect for active sports. This is not surprising, as relaxing the normality assumption does not come without cost in precision. As for the individual characteristics and contextual peer effects, I notice that they are fairly similar to the ones obtained earlier.

Based on the evidence presented in this section, it appears that the results are in favour of the presence of encouraging effects of friends in exercising. Exercising seems to be *better with a buddy* (Jago et al., 2010).

[30]Note that the fast food social multiplier estimated in Fortin and Yazbeck (2011) is larger than the physical activity multiplier estimated in this paper. The same rational applies: it is more appealing to eat today fast food today and stop tomorrow.

3.5.4 Health production function estimates

The estimates of the peer effect model are consistent with the hypothesis that having a friend who exercises might help in increasing one's own exercising frequency. The presence of peer effects in exercising is valuable because of its positive impacts on health outcomes. In this section, I report the results of the production function presented earlier in an attempt to provide evidence on the presence of a relationship between physical activity and self-rated health status. The dependant variable used is self-reported general health status. The lagged value of self-reported health status and lagged values of physical activity as well as the instruments previously utilized in the peer effect model were used as instruments.

Estimation results are reported in table 3.8 are in line with my expectations. Past health status is positively related to current health status. Being in good health today is significantly explained by past healthy status. This suggests that the effects of an exogenous shock on physical activity will have a stronger effect in the long run. Also, physical activity has a positive and significant effect on self-reported health. The presence of a causal link between physical activity and self-reported health status is compatible with the results obtained by Pisinger et al. (2009). It is also, generally compatible with the medical and epidemiological literature on the impacts of physical activity and health outcomes (see section 2).

The consistency of the production function estimates depend on the validity of the instruments used. I test the validity of the instruments using a Sargan test and report the statistic in table 3.8. I do not reject the hypothesis that the instruments are valid and that the over-identification restrictions are satisfied. In addition, given that I do not reject the validity of the instruments, the Sargan test provides an indirect test for the presence serial autocorrelation in the idiosyncratic term ν_{it}.[31]

Finally, to highlight the presence of a transmission mechanism between physical activity and health, I combine results from the production function with those obtained in the peer effect. This allows me to compute the short run and long run impacts of an exogenous shock on physical activity on health status via the social multiplier. The obtained result suggests that, in the short run, exercising one extra time improves health status by 0.10 (=1.12 $\times$ 0.09) or by approximately by 5%. In the long run this impact becomes 0.13 (=1.12 $\times$ 0.09 $\times$ $\frac{1}{1-0.25}$). This allows for the conclusion that exercising impact on physical health can be amplified by the presence of social multiplier.

[31]The lagged variables will not be valid as instruments in the presence of serial autocorrelation. Note that the length of the panel is too short (i.e., T=3 in my case) to perform a direct test for serial correlation

3.6 Conclusion

In this paper, I investigate the presence of peer effects in physical activity and the extent to which these effects can be considered as a mechanism through which peer effects in health outcomes flow. To address this question, I first estimate a peer effect model to examine the nature and the size of the peer effects in physical activity. I then augment the peer effect model with a health production function. Results from the peer effect model show that there are peer effect in exercising: the associated multiplier effect is 1.12. This shows that a friend's mean exercising frequency increases an own adolescent exercising frequency, suggesting that an individual lifestyle is affected by his peer's lifestyle. As for recreational sports or active sports, estimation results are not in favour of the presence of a social multiplier for these types of physical activity. It appears that most of the endogenous effect associated for these activities can be attributed to the presence of common preferences for certain sports type. It is not hard to conceive that people who perform outdoors activities (i.e., recreational activities) may have a common preference for exercising outdoors. A similar rational can be applied for active sports which is composed of team sports only.

In addition to the peer effect model, I estimate a dynamic health production function where I relate an individual's health status to his physical activities. The results show that exercising has a positive significant impact on health of 0.09 (i.e., 4 %). This impact increases to 0.10 (i.e., 5 %) in the short run and 0.13 (i.e., 6%) in the long-run once I account for the social multiplier. This means that public policies can exploit the social multiplier to increase the spread of exercising through tax (e.g., gas tax) and subsidy reforms encouraging physical activity.

Recently, there is a rising interest in the presence of peer effects in health outcomes and individual well being and happiness in general. One could extend this paper to investigate the presence of such peer effects and analyze the potential underlying mechanisms that could mitigate them. A study by Fowler and Christakis (2008a) shows that happiness spreads through social network suggesting that an individual happiness affect the happiness of those who are in his reference group however the underlying mechanism is not identified. Further investigation in this direction could be interesting. Integrating health indices is another possible extension to this paper. Instead of using self-reported health status, one can construct a health related quality of life index. To my knowledge no work has been done in this direction. On the methodological level, a possible extension can be to turn to non linear ordered models (e.g., logit) to estimate peer effects and augment it with an ordered production health production function. Finally, one can take the the two step link formation approach a step further by finding instrumental variables that directly affect the formation of social networks but does not affect the outcomes of interest. To my knowledge no work has been done in this direction yet.

Table 3.1: Descriptive Statistics for the Peer Effect Model

	Mean	**S.D**
Exercise	1.6298	1.0066
Recreational Sports	1.4347	1.1366
Active Sports	0.5861	0.8777
Female	0.5032	0.5001
Age	16.3477	1.4516
White	0.5751	0.4945
Black	0.1410	0.3482
Asian	0.0092	0.0956
Native	0.1290	0.3352
Other	0.1457	0.3529
Mother Present	0.8641	0.3427
Mother No High School	0.1399	0.3470
Mother High School	0.3638	0.4812
Mother Some College	0.1929	0.3947
Mother College	0.1940	0.3956
Mother Advanced	0.0656	0.2477
Mother Don't Know	0.0438	0.2046
Father No High School	0.1387	0.3458
Father High School	0.3253	0.4686
Father Some college	0.1698	0.3756
Father College	0.1911	0.3933
Father Advanced	0.0794	0.2705
Father Don't Know	0.0760	0.2651
Father Missing	0.0196	0.1386
Grade 7-8	0.1134	0.3172
Grade 9-10	0.2689	0.4435
Grade 11-12	0.6177	0.4861
Allowance per week	8.4116	11.7141
Observations		1737

Table 3.2: Descriptives Statistics for the Production Function

	Mean	**S.D**
Self Reported General Physical Health		
Wave 1	2.0968	0.8935
Wave 2	2.1024	0.8731
Wave 3	1.9819	0.8698
Observations		1436

Table 3.3: Peer Effects in Physical Activity GSAR- OLS

	Exercising			Active Sports			Recreational Sports		
	Coef.		**S.E**	**Coef.**		**S.E**	**Coef.**		**S.E**
Endogenous Peer Effects	0.1669	***	0.0398	0.2445	***	0.0392	0.1965	***	0.0387
Individual Characteristics									
Female	-0.0550		0.0528	-0.6673	***	0.0551	-0.2222	***	0.0453
Age	-0.0895	***	0.0304	-0.1586	***	0.0318	-0.0047		0.0261
White	0.1319	**	0.0643	0.0427		0.0671	0.0926	*	0.0552
Mother Present	-0.0683		0.0711	0.0581		0.0743	0.0567		0.0610
Mother No HIgh School (Omitted)									
Mother High School	0.0615		0.0855	-0.0425		0.0891	-0.0391		0.0733
Mother Some College	0.0687		0.0949	-0.1057		0.0990	-0.0987		0.0814
Mother College	0.1257		0.0987	-0.0419		0.1030	-0.0086		0.0848
Mother Advanced	0.0685		0.1305	0.0021		0.1363	0.1421		0.1119
Mother Don't Know	0.0787		0.1454	-0.3037	**	0.1516	-0.1407		0.1246
Father No High School (Omitted)									
Father High School	-0.1070		0.0850	0.1025		0.0887	0.0961		0.0730
Father Some College	-0.0449		0.0958	0.1551		0.1000	0.0711		0.0823
Father College	0.0281		0.0977	0.1690	*	0.1019	0.0528		0.0838
Father Advanced	-0.0908		0.1227	-0.0388		0.1280	0.0994		0.1052
Father Don't Know	-0.0403		0.1195	0.0645		0.1247	0.1528		0.1025
Father Missing	0.1798		0.1842	-0.0085		0.1922	0.0169		0.1580
Grade 7-8 (Omitted)									
Grade 9-10	0.2542	**	0.1280	0.2214	*	0.1333	-0.3734	***	0.1096
Grade 11-12	0.0716		0.1529	0.1152		0.1595	-0.5274	***	0.1312
Allowance	0.0033		0.0022	0.0048	**	0.0023	0.0026		0.0019
N=1721									

continued on next page

Table 3.3: Continued

	Exercising			Active Sports			Recreational Sports		
	Coef.		**S.E**	**Coef.**		**S.E**	**Coef.**		**S.E**
Contextual Peer Effects									
Female	0.0804		0.0864	0.1355		0.0946	0.0619		0.0743
Age	-0.0058		0.0151	-0.0362	**	0.0158	-0.0207	*	0.0129
White	-0.1720	*	0.0891	0.0827		0.0933	0.0293		0.0765
Mother Present	-0.0486		0.1186	-0.0753		0.1243	0.0906		0.1016
Mother No HIgh School (Omitted)									
Mother High School	-0.0280		0.1377	0.3805	***	0.1433	-0.0321		0.1179
Mother Some College	0.0482		0.1537	0.3209	**	0.1598	0.0108		0.1315
Mother College	0.0580		0.1594	0.1017		0.1658	-0.1356		0.1363
Mother Advanced	0.4113	*	0.2099	0.3479		0.2191	-0.1839		0.1801
Mother Don't Know	-0.1226		0.2616	0.4143		0.2735	-0.2030		0.2245
Father No High School (Omitted)									
Father High School	-0.1827		0.1408	-0.1863		0.1472	-0.1449		0.1207
Father Some College	-0.0802		0.1549	0.0258		0.1619	-0.0533		0.1329
Father College	-0.3059	*	0.1611	0.0105		0.1690	-0.0815		0.1382
Father Advanced	-0.2056		0.1940	0.1625		0.2024	0.1377		0.1665
Father Don't Know	-0.1830		0.2021	0.1352		0.2110	-0.0693		0.1734
Father Missing	0.0550		0.4012	-0.3169		0.4185	-0.1387		0.3456
Grade 7-8 (Omitted)									
Grade 9-10	-0.1527		0.1581	-0.0076		0.1651	0.1207		0.1363
Grade 11-12	0.0337		0.1756	0.1393		0.1845	0.2066		0.1520
Allowance	-0.0013		0.0038	-0.0016		0.0039	-0.0017		0.0032
Constant	3.0134	***	0.4603	3.9897	***	0.4811	1.1763	***	0.3943
N=1721									

***Significant at 1% level **Significant at 5% level *Significant at 10% level

Table 3.4: Peer Effects in Physical Activity GSAR- OLS Within

	Exercising			Active Sports			Recreational Sports		
	Coef.		**S.E**	**Coef.**		**S.E**	**Coef.**		**S.E**
Endogenous Peer Effects	0.1193	***	0.0417	0.2113	***	0.0414	0.1578	***	0.0406
Individual Characteristics									
Female	-0.0524		0.0529	-0.6630	***	0.0555	-0.2281	***	0.0453
Age	-0.0833	***	0.0324	-0.1500	***	0.0340	-0.0049		0.0277
White	0.1594	**	0.0809	-0.0240		0.0848	0.1714	**	0.0691
Mother Present	-0.0600		0.0711	0.0723		0.0745	-0.0706		0.0607
Mother No HIgh School (Omitted)									
Mother High School	0.0500		0.0862	-0.0656		0.0904	0.0030		0.0737
Mother Some College	0.0609		0.0958	-0.1330		0.1005	-0.0632		0.0819
Mother College	0.0973		0.0995	-0.0612		0.1043	0.0205		0.0851
Mother Advanced	0.0448		0.1312	-0.0212		0.1377	0.1644		0.1121
Mother Don't Know	0.0876		0.1457	-0.3081	**	0.1527	-0.1223		0.1245
Father No High School (Omitted)									
Father High School	-0.1114		0.0855	0.0817		0.0897	0.0928		0.0731
Father Some College	-0.0182		0.0968	0.1410		0.1015	0.0725		0.0827
Father College	0.0522		0.0987	0.1645		0.1034	0.0609		0.0843
Father Advanced	-0.0556		0.1241	-0.0523		0.1301	0.1272		0.1061
Father Don't Know	-0.0310		0.1195	0.0665		0.1253	0.1460		0.1021
Father Missing	0.1525		0.1850	-0.0241		0.1940	-0.0300		0.1581
Grade 7-8 (Omitted)									
Grade 9-10	0.2815	*	0.1494	0.2828	*	0.1566	-0.4074	***	0.1277
Grade 11-12	0.0892		0.1763	0.2521		0.1848	-0.6995	***	0.1507
Allowance	0.0031		0.0022	0.0055	**	0.0023	0.0028		0.0019
N=1721									

continued on next page

Table 3.4: Continued

	Exercising		Active Sports		Recreational Sports	
	Coef.	**S.E**	**Coef.**	**S.E**	**Coef.**	**S.E**
Contextual Peer Effects						
Female	0.1079	0.0896	0.1339	0.0989	0.0825	0.0766
Age	0.0395	0.0540	-0.0054	0.0566	-0.0324	0.0460
White	-0.2277 *	0.1232	0.0353	0.1292	0.0456	0.1053
Mother Present	-0.0092	0.1219	-0.0298	0.1280	0.0152	0.1041
Mother No HIgh School (Omitted)						
Mother High School	-0.0062	0.1430	0.3478 **	0.1498	0.0147	0.1222
Mother Some College	0.0817	0.1587	0.3163 *	0.1665	0.0795	0.1355
Mother College	0.0926	0.1638	0.1344	0.1715	-0.0546	0.1398
Mother Advanced	0.3891 *	0.2115	0.3496	0.2218	-0.1687	0.1808
Mother Don't Know	0.0273	0.2647	0.3998	0.2782	-0.0806	0.2262
Father No High School (Omitted)						
Father High School	-0.1836	0.1414	-0.1969	0.1485	-0.1857	0.1208
Father Some College	-0.0432	0.1558	0.0262	0.1635	-0.1058	0.1331
Father College	-0.2135	0.1650	0.0089	0.1737	-0.0730	0.1410
Father Advanced	-0.1015	0.1995	0.1247	0.2092	0.1750	0.1705
Father Don't Know	-0.0482	0.2021	0.0961	0.2119	-0.1344	0.1728
Father Missing	0.3267	0.4073	-0.1584	0.4269	-0.3696	0.3496
Grade 7-8 (Omitted)						
Grade 9-10	-0.1790	0.1954	-0.1007	0.2045	0.1373	0.1675
Grade 11-12	-0.0674	0.2445	-0.0606	0.2563	0.3530 *	0.2107
Allowance	-0.0012	0.0039	-0.0003	0.0041	-0.0001	0.0033
N=1721						

***Significant at 1% level **Significant at 5% level *Significant at 10% level

Table 3.5: Peer Effects in Physical Activity GSARAR- MLE

	Exercising			Active Sports			Recreational Sports		
	Coef.		**S.E**	**Coef.**		**S.E**	**Coef.**		**S.E**
Endogenous Peer Effects	0.0545		0.1190	-0.0583		0.0648	-0.1078		0.0816
Auto Correlation	0.0549		0.1171	0.2227	***	0.0524	0.2481	***	0.0749
Individual Characteristics									
Female	-0.0517		0.0522	-0.6649	***	0.0543	-0.2348	***	0.0442
Age	-0.0793	***	0.0311	-0.1460	***	0.0315	-0.0107		0.0252
White	0.1544	*	0.0793	-0.0313		0.0834	0.1713	**	0.0680
Mother Present	-0.0614		0.0705	0.0583		0.0741	-0.0669		0.0604
Mother No HIgh School (Omitted)									
Mother High School	0.0570		0.0849	-0.0475		0.0891	-0.0080		0.0726
Mother Some College	0.0691		0.0944	-0.1157		0.0990	-0.0672		0.0807
Mother College	0.1120		0.0980	-0.0558		0.1028	0.0116		0.0838
Mother Advanced	0.0572		0.1295	-0.0133		0.1358	0.1598		0.1107
Mother Don't Know	0.0869		0.1440	-0.2995	**	0.1514	-0.1296		0.1234
Father No High School (Omitted)									
Father High School	-0.1101		0.0844	0.0654		0.0886	0.0894		0.0722
Father Some College	-0.0182		0.0955	0.1397		0.1004	0.0681		0.0818
Father College	0.0602		0.0976	0.1567		0.1024	0.0620		0.0835
Father Advanced	-0.0530		0.1227	-0.0587		0.1289	0.1331		0.1052
Father Don't Know	-0.0280		0.1182	0.0768		0.1243	0.1355		0.1014
Father Missing	0.1692		0.1829	-0.0362		0.1920	-0.0508		0.1565
Grade 7-8 (Omitted)									
Grade 9-10	0.2891	**	0.1458	0.2904	*	0.1507	-0.3878	***	0.1225
Grade 11-12	0.0964		0.1696	0.2337		0.1752	-0.6374	***	0.1422
Allowance	0.0031		0.0022	0.0057	**	0.0023	0.0025		0.0019
N= 1737									

continued on next page

Table 3.5: Continued

	Exercising		Active Sports		Recreational Sports	
	Coef.	**S.E**	**Coef.**	**S.E**	**Coef.**	**S.E**
Contextual Peer Effects						
Female	0.0822	0.0875	-0.0683	0.0923	0.0373	0.0755
Age	-0.0073	0.0157	-0.0623 ***	0.0190	-0.0296 *	0.0162
White	-0.1625 *	0.0981	-0.0305	0.1094	0.1294	0.0897
Mother Present	-0.0506	0.1188	-0.0495	0.1269	-0.0175	0.1036
Mother No HIgh School (Omitted)						
Mother High School	-0.0017	0.1379	0.3291 **	0.1459	0.0481	0.1191
Mother Some College	0.0803	0.1532	0.2899 *	0.1628	0.0980	0.1329
Mother College	0.0948	0.1586	0.0945	0.1684	0.0049	0.1375
Mother Advanced	0.4030 *	0.2094	0.3017	0.2229	-0.0886	0.1820
Mother Don't Know	-0.0116	0.2605	0.2918	0.2723	-0.2146	0.2218
Father No High School (Omitted)						
Father High School	-0.1741	0.1396	-0.1679	0.1477	-0.1324	0.1203
Father Some College	-0.0520	0.1543	0.0835	0.1642	-0.0633	0.1341
Father College	-0.2238	0.1625	0.0271	0.1737	-0.1021	0.1420
Father Advanced	-0.1234	0.1963	0.1125	0.2088	0.1959	0.1704
Father Don't Know	-0.0614	0.1997	0.1399	0.2091	-0.1057	0.1700
Father Missing	0.3895	0.4015	-0.0922	0.4169	-0.3431	0.3389
Grade 7-8 (Omitted)						
Grade 9-10	-0.1488	0.1855	-0.0089	0.2008	0.0081	0.1635
Grade 11-12	0.0166	0.1999	0.0722	0.2200	0.0850	0.1778
Allowance	-0.0016	0.0038	0.0014	0.0041	-0.0011	0.0033
N= 1737						
Log Likelihood	-2401.4862		-2482.3320		-2131.3543	
ex						

*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

Table 3.6: Peer Effects in Physical Activity GSAR-MLE

	Exercising			Active Sports			Recreational Sports		
	Coef.		**S.E**	**Coef.**		**S.E**	**Coef.**		**S.E**
Endogenous Peer Effects	0.1076	***	0.0345	0.1691	***	0.0339	0.1337	***	0.0341
Individual Characteristics									
Female	-0.0517		0.0523	-0.6609	***	0.0548	-0.2303	***	0.0447
Age	-0.0792	***	0.0311	-0.1460	***	0.0326	-0.0069		0.0266
White	0.1561	**	0.0794	-0.0244		0.0833	0.1682	**	0.0679
Mother Present	-0.0611		0.0704	0.0662		0.0739	-0.0606		0.0603
Mother No HIgh School (Omitted)									
Mother High School	0.0566		0.0850	-0.0573		0.0891	-0.0024		0.0727
Mother Some College	0.0682		0.0945	-0.1253		0.0990	-0.0672		0.0808
Mother College	0.1112		0.0980	-0.0553		0.1028	0.0165		0.0838
Mother Advanced	0.0553		0.1296	-0.0184		0.1359	0.1710		0.1109
Mother Don't Know	0.0868		0.1439	-0.3054	**	0.1509	-0.1180		0.1231
Father No High School (Omitted)									
Father High School	-0.1084		0.0844	0.0764		0.0885	0.0959		0.0722
Father Some College	-0.0179		0.0955	0.1409		0.1002	0.0761		0.0817
Father College	0.0614		0.0976	0.1511		0.1023	0.0668		0.0835
Father Advanced	-0.0529		0.1226	-0.0542		0.1286	0.1337		0.1049
Father Don't Know	-0.0278		0.1181	0.0756		0.1238	0.1417		0.1010
Father Missing	0.1705		0.1827	-0.0270		0.1916	-0.0366		0.1563
Grade 7-8 (Omitted)									
Grade 9-10	0.2843	*	0.1463	0.2939	*	0.1534	-0.3777	***	0.1251
Grade 11-12	0.0941		0.1701	0.2377		0.1783	-0.6325	***	0.1455
Allowance	0.0031		0.0022	0.0057	**	0.0023	0.0026		0.0019
N= 1737									

continued on next page

Table 3.6: Continued

	Exercising		Active Sports		Recreational Sports	
	Coef.	**S.E**	**Coef.**	**S.E**	**Coef.**	**S.E**
Contextual Peer Effects						
Female	0.0804	0.0869	0.0983	0.0902	0.0906	0.0736
Age	-0.0028	0.0153	-0.0233	0.0161	-0.0182	0.0134
White	-0.1612 *	0.0965	0.0096	0.1011	0.0690	0.0819
Mother Present	-0.0434	0.1182	-0.0240	0.1240	0.0328	0.1011
Mother No HIgh School (Omitted)						
Mother High School	-0.0084	0.1376	0.3274 **	0.1443	0.0618	0.1177
Mother Some College	0.0772	0.1526	0.2834 *	0.1600	0.0907	0.1305
Mother College	0.0853	0.1580	0.1049	0.1658	-0.0214	0.1352
Mother Advanced	0.3948 *	0.2084	0.3354	0.2185	-0.1194	0.1782
Mother Don't Know	-0.0309	0.2605	0.3798	0.2732	-0.0979	0.2228
Father No High School (Omitted)						
Father High School	-0.1681	0.1392	-0.1794	0.1461	-0.1838	0.1191
Father Some College	-0.0497	0.1536	0.0290	0.1610	-0.1106	0.1313
Father College	-0.2244	0.1616	0.0233	0.1694	-0.1041	0.1382
Father Advanced	-0.1205	0.1954	0.1371	0.2049	0.1470	0.1670
Father Don't Know	-0.0568	0.1997	0.0974	0.2093	-0.1514	0.1706
Father Missing	0.3738	0.4019	-0.1582	0.4215	-0.4009	0.3439
Grade 7-8 (Omitted)						
Grade 9-10	-0.1510	0.1840	-0.0831	0.1930	0.0585	0.1566
Grade 11-12	0.0127	0.1974	0.0124	0.2070	0.1767	0.1667
Allowance	-0.0017	0.0038	0.0003	0.0040	-0.0008	0.0032
N= 1737						
Log Likelihood	-2401.5758		-2483.9386		-2133.0446	

*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

Table 3.7: Peer Effects in Physical Activity GSAR- GS2SLS

	Exercising			Active Sports			Recreational Sports		
	Coef.		**S.E**	**Coef.**		**S.E**	**Coef.**		**S.E**
Endogenous Peer Effects	0.0354		0.4780	-0.4736		1.7391	-0.0150		0.3063
Individual Characteristics									
Female	-0.0505		0.0541	-0.6831	***	0.0786	-0.2328		0.0462
Age	-0.0824	***	0.0329	-0.1615	***	0.0468	-0.0059		0.0279
White	0.1580	*	0.0813	-0.0371		0.0973	0.1654	**	0.0703
Mother Present	-0.0578		0.0722	0.1230		0.1517	-0.0703		0.0610
Mother No HIgh School (Omitted)									
Mother High School	0.0431		0.0949	0.0028		0.1991	0.0131		0.0762
Mother Some College	0.0590		0.0965	-0.0951		0.1448	-0.0559		0.0833
Mother College	0.0983		0.0997	-0.0057		0.1801	0.0333		0.0884
Mother Advanced	0.0512		0.1363	0.1023		0.3470	0.1678		0.1128
Mother Don't Know	0.0825		0.1487	-0.3241		0.1696	-0.1226		0.1251
Father No High School (Omitted)									
Father High School	-0.1111		0.0856	0.0693		0.1016	0.0899		0.0736
Father Some College	-0.0166		0.0973	0.1096		0.1352	0.0693		0.0833
Father College	0.0467		0.1036	0.1589		0.1124	0.0507		0.0866
Father Advanced	-0.0618		0.1292	-0.0401		0.1436	0.1336		0.1072
Father Don't Know	-0.0298		0.1198	0.1079		0.1712	0.1490		0.1028
Father Missing	0.1529		0.1852	-0.0593		0.2274	-0.0330		0.1590
Grade 7-8 (Omitted)									
Grade 9-10	0.2763	*	0.1524	0.3326		0.2109	-0.4035	***	0.1285
Grade 11-12	0.0902		0.1765	0.2438		0.2003	-0.6927	***	0.1519
Allowance	0.0031		0.0022	0.0056	**	0.0025	0.0026		0.0019
N=1721									

continued on next page

Table 3.7: Continued

	Exercising		Active Sports		Recreational Sports	
	Coef.	**S.E**	**Coef.**	**S.E**	**Coef.**	**S.E**
Contextual Peer Effects						
Female	0.0985	0.1041	-0.3820	1.3140	0.0660	0.0822
Age	0.0306	0.0739	-0.0907	0.2252	-0.0228	0.0492
White	-0.2223 *	0.1272	-0.0361	0.2286	0.0589	0.1084
Mother Present	-0.0015	0.1296	0.1250	0.4167	0.0125	0.1047
Mother No HIgh School (Omitted)						
Mother High School	0.0089	0.1670	0.2824	0.2316	0.0367	0.1287
Mother Some College	0.1006	0.1918	0.1388	0.4850	0.0894	0.1372
Mother College	0.1117	0.1966	0.0642	0.2568	-0.0347	0.1448
Mother Advanced	0.3966 *	0.2159	0.4406	0.3325	-0.1510	0.1843
Mother Don't Know	0.0457	0.2849	0.0304	0.9843	-0.0605	0.2300
Father No High School (Omitted)						
Father High School	-0.1926	0.1503	-0.0235	0.4686	-0.1957	0.1227
Father Some College	-0.0419	0.1561	0.1521	0.3651	-0.1042	0.1339
Father College	-0.2091	0.1671	0.2711	0.6916	-0.0747	0.1418
Father Advanced	-0.1002	0.1998	0.0411	0.3097	0.1761	0.1714
Father Don't Know	-0.0423	0.2051	0.1790	0.3106	-0.1208	0.1753
Father Missing	0.3054	0.4251	-0.0282	0.5665	-0.5133	0.4327
Grade 7-8 (Omitted)						
Grade 9-10	-0.1571	0.2316	-0.0749	0.2300	0.0701	0.2056
Grade 11-12	-0.0572	0.2516	0.0114	0.3313	0.2364	0.2947
Allowance	-0.0009	0.0043	0.0062	0.0169	0.0007	0.0037
N= 1721						

***Significant at 1% level **Significant at 5% level *Significant at 10% level

Table 3.8: Physical Health Production Function

	Coef	**S.E**	
Constant	0.1354	0.0553	***
Δ General Health$_{t-1}$	0.2542	0.0623	***
Δ Recreational Sports	0.0782	0.0494	
Δ Active Sports	0.0447	0.0416	
Δ Exercising	0.0927	0.0409	**
Sargan Chi2 test(54)		56.5290	
N		**1436**	

Chapter 4

Isolating the Effect of the South African Old-Age Pension on Grandchildren's Health: Role of Selection Bias

4.1 Introduction

A large number of studies are concerned with measuring the impact of income on health.[1] Some of these focus on the effect of income on children's health (Case et al., 2002; Currie and Stabile, 2003) and others on the impact of conditional cash transfers on children's health (for a complete review see Fiszbein et al., 2009). Nevertheless, there are very few studies on the channels through which family income may affect children's health. Recently, a finer line of research looked at the mechanisms through which income may affect children health outcomes.[2] Propper et al. (2007), for instance, find little evidence of a direct impact of income on children's health. Rather, their study suggests that improvements in maternal health is the channel through which income flows to children.[3] This paper contributes to this existing literature by investigating the mechanism through which the South African old-age pension program has produced its effects on children's health and by shedding some light on the channel through which the pension program may have affected children's health.

The use of cash transfers programs as poverty alleviation mean in developing coun-

[1] See for instance Adler et al. (1994), Van Doorslaer et al. (1997),Deaton and Paxton (1998) and Marmot and Wilkinson (2006).

[2] See for instance Currie et al. (2007),Propper et al. (2007) and Khanam et al. (2009).

[3] Similar findings are found in a recent study by Khanam et al. (2009).

tries is very common. The South African old-age pension is targeted to provide support for the economically deprived elderly population. It is a sizeable amount of cash that represents more than double the median per capita income for a Black household. Since 1991, the Black South African population gained a color-blind access to this program which resulted in a drastic increase in the financial resources available to this population. This pension is generous enough to produce unintended effects on living arrangements (Edmonds et al., 2005). In fact, the program may have prompted a grandparent and his adult children to form an extended household. This, in turn, may have produced an impact on children's health by allowing grandparents to provide better care to grand-children.[4] Viewed from this perspective, the pension may have produced its impact on children's health directly through an income effect (e.g., food of better quality) or indirectly through its impact on living arrangements (e.g., more time allocated to the production of child care) or a combination of both. Our objective is to provide a better understanding on how this increase in income has affected children's health.

In this paper, we investigate living arrangements as a possible channel through which income impacts children's health. Unlike much of the literature on the mechanism of income-health gradient, we study the case of a developing country and exploit the unique nature of the South African old age pension quasi-experiment to identify the impact of the this pension program on children's health (as in Case, 2001; Duflo, 2003). In theory, there are different motivations behind the formation of extended households.[5] We provide a discussion of each of these reasons and frame their possible impacts namely: sorting of children according to their health status. In doing so, we show that most of the reasons for which extended household are formed cannot be proved to be independent of children's health status. This suggests that further understanding of how this pension produced its impacts is needed.

While our focus on household structure as a possible mechanism is motivated by the nature of living arrangements in South Africa, our interest in the direct impact of income on health and its transmission mechanisms is policy driven. In fact, the family environment in early childhood is a key determinant for a child's success in adult life. These effects can mediated through the child's physical as well as cognitive health. According to Heckman (2006) living in a disadvantaged environment at an early age has a detrimental effect on adult age success. He also argues that targeting interventions towards disadvantaged children at an early age (i.e., preventive measures) is more efficient than at later stages of life (i.e., remedying). Therefore, if changes in household structure are an important channel through which the South African old age pension program affects children's health at an early age, then this pension program may be more effective in addressing children's wellbeing if coupled with interventions that facilitate changes in living arrangements.

[4]Better care can be either through more time invested in the production of child health or better surveillance of resources allocated to children's health.

[5]Extended family bringing together grandparents, adult children and grand-children is actually a common living arrangement in South Africa.

Estimating the impact of income on health presents serious identification challenges.[6] First, the correlation between income and health may reflect the effect of unobserved factors on both outcomes. In this case, better data helps. But even in the absence of spurious correlation, identification is hindered if one's health and one's income are simultaneously determined.[7] Second, the pension's unintended impacts on household structure (i.e., the emergence of extended living arrangement) makes it difficult to disentangle the impacts of pension itself from the impact of the changes in the household structure resulting from pension eligibility (Edmonds et al., 2005). Further, if extended living arrangements' decision is based on children's health status at birth then identifying the real impact of the pension may be hindered further through selection bias. A similar problem was pointed out by Angrist et al. (2002) and Angrist et al. (2006) for estimating the effect of school vouchers on test scores, when school vouchers affects the composition of the pool of students enrolled in school.

In order to address the first issue, we rely on quasi-experimental nature of the south african old age pension that exogenously manipulates family income. In this case, the effect of income on health can be readily estimated from a simple difference in means between treated and control units. However, even such an ideal setting is not sufficient to understand the channels through which income may affect health (second issue). To address this possibility, we draw upon the literature of treatment effects (Flores and Flores-Lagunes, 2007) by treating living arrangements as a post-treatment variable. We provide an econometric framework to identify the impact of income on children's health which is not due to changes in living arrangements, *i.e.* a *net income effect.* Based on the sample of children under the age of six present in extended households, we exploit the timing of effects and the longitudinal aspect of the data to block the impact of the treatment on the post-treatment variable and circumvent the selection issue. We do not expect household formation to be endogenous to child's health status when the child is born in a household where the elderly relative is already present. Thus by conditioning on the presence of the elderly presence prior the the child's birth, we remove the selection bias resulting from sorting by blocking the impact of the pension on household structure.

Unlike Duflo (2003), the results show that when we condition on household composition and block the effect of the pension on living arrangements, we do not find a significant effect on child's anthropometric measures, suggesting that the effect that was found was entirely due to the selection effect. Such a selection effect is consistent with elderly pairing up with children of better health status under the program. Our result suggests that the effect found by Duflo may be entirely due to changes in the composition of households with no improvements in children's health. It suggests that pension cash transfers targeted to the elderly poor do not trickle down to children. It cautions against conditioning on variables that are themselves affected by a treatment

[6]See for instance Adams et al. (2003), Meer et al. (2003), Contoyannis et al. (2004), Frijters et al. (2005) and Lindahl (2005).

[7]To cite Deaton (2002):"Income might cause health, health might cause income or both might be correlated with other factors."

when estimating an average treatment effect on beneficiaries.

The remaining of the paper unfolds as follows. In section 2, we present a simple theoretical framework to frame the expected impacts of the South African Old Age Pension. In Section 3 we present our identification and empirical strategy. In section 4 we provide a brief description of the South African old age pension program and of the Kwazulu-Natal Income Dynamics Study data. In section 5 we present our empirical results. Finally, section 6 will conclude.

4.2 Background on The South African Old Age Pension

The South African population is of three types : Whites, Blacks and Indians. Prior to 1991, only whites had full access to the pension. In 1991, the program became color-blind and the Black population gained an access to this program.[8] This led to a drastic increase in the financial resources available to the Black South African population. The pension represents a sizeable amount of cash. It represents more than double the median per capita income (for a Black household) and is the principal source of income after earnings gained from paid jobs and informal work (Moller and Ferreira, 2003). There are two eligibility criteria: means and age. Women are eligible at the age of 60 whereas men are eligible at the age of 65. Given that Black elderly people are among the poorest in South Africa, almost all of them are means-eligible. This thus makes age the principal binding eligibility criterion. In particular, pension income does not depend on previous work history, the income of household members. Individual can collect their pension regardless of labor market attachment.

Pension money is paid in mobile points such as schools, civic centers and police stations which makes accessible for most pension eligibiles. Pension income is typically pooled with income of other household members and shared with the extended family. According to the Non-Contributory Pensions and Poverty Study Survey, on average 76% of Black South Africans pension recipients report that none of the pension money was for their own use. Qualitative evidence also suggest that the elderly provide an important share of child care in extended families (Burman, 1995). Taking a historical perspective, Burman (1996) documents the causes of the reversal of the typical flow of support from the younger to the older generation that Black South African families have experienced. He points to the effect of the restriction of Black South Africans to townships and "homelands" during the apartheid era on Black South African living arrangements. Men were often allowed to temporary work in certain restricted areas. But they were generally prohibited to bring their spouses and children along. In this context, mothers started to rely on the support provided by other family members,

[8]This pension program is financed by general tax revenues that are collected on a national basis

and in particular grandparents. This pattern was reinforced by migration of prime age adult, driven by poverty and by the increase in opportunities for Blacks after the end of the apartheid era; see Anderson (2000), Case and Deaton (1998).

4.3 Theoretical Discussion

The objective of this section is to explain how the introduction of the South African old-age pension may affect children's living arrangements and health outcomes. We distinguish between a treatment effect on health and a selection effect on health. We relate the selection effect to the mechanism driving the effect of the program on living arrangements. We also explore the mechanisms that could give rise to the treatment effect of the pension on health.

4.3.1 Impact of the old-age pension on living arrangements

Extended households are typically three generation households in which children, their mother/father, and at least one elderly co-reside. Skip generation household, where mother and father are absent and the elderly co-reside with children, are also considered as extended. We contrast the extended structure to the nuclear structure. Elderly who are not co-residing with children are considered to be living in a nuclear household. Similarly, we consider that children who are not co-residing with an elderly live in nuclear households.

We follow Edmonds et al. (2005) and distinguish between three explanations for why the introduction of the old-age pension may affect living arrangements. First, the elderly propensity to consume his/her preferred living arrangements may increase. If both the elderly and the children's parents prefer co-residing together than living in separate households, then a new extended household is formed. The opposite holds if their preference is towards more privacy. If an extended household is preferred by the elderly but not by the children's parents, a new extended household may still be formed if the elderly can compensate the children's parents for their loss in terms of private life.

Second, even in the absence of heterogeneity over preferences for living arrangements, a new extended household may be formed if the elderly finds that this arrangement eases the enforcement of contracts. Living in an extended household may, for instance, help the elderly monitoring the behaviour of his adult children in exchange for his financial support, in which case the pension program may result in the formation of extended households (Pollak, 1988). In contrast, the pension program, by providing public social insurance, may crowd-out informal risk-sharing arrangements and leads

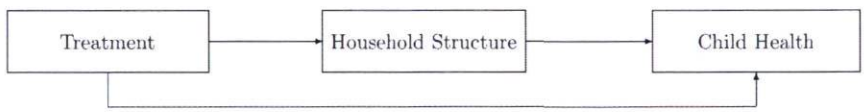
extended households to split.

Finally, in the absence of heterogeneity over preferences for living arrangements and in the absence of issues of enforcement of intra-family contracts, a new extended household may be formed if this arrangement allows the family to increase household production efficiency. This may, for instance, be the case if by taking care of the children, the elderly allows the adult child to increase his/her productivity on the labor market.

4.3.2 Selection effect, treatment effect or both?

Children living with pensioners may be found to have better health outcomes than children living with non-beneficiary elderly (as in Duflo 2003). The effect can be entirely driven by selection; it can be due to a direct income effect; or it may reflect both a selection and a treatment effect. The causal effect of the treatment may arise through the change in household structure and/or from the extra income made available to the elderly.

Figure 4.1: Treatment Effect on Child Health



An effect on children's health outcomes may be obtained via the change in household structure (see figure 4.1). The elderly may, for instance, have more say over the child's health from the fact that they now co-reside. If she values the child's health more than the child's parent, then child's health may be improved. This causal effect related to the change in household structure may also arise from the gain in efficiency in household production. The gain in efficiency may result in children being better off in terms of health outcomes if, for instance, more time is spent on child care. As discussed previously, the change in household structure may also be motivated by the change in family contractual arrangements. This change may, in turn, affect child's health.

Conditional on living in an extended household, children's health may also improve as a result of the extra income provided by the pension program (again, see figure 4.1). Their health may improve because the extra income is spent in doctor's visits and medicine. It may further increase if the elderly say over the household's expenditure pattern is greater as a result of the program and she has a strong preference for spending the extra income on inputs to child's health. This extra income may also allow the household to allocate more time to child care.

The pension program may not only result in more elderly and children co-residing, it may also change the composition of the pool of children living with elderly. Extended households formed as a result of the program may foster children with better/worse health outcomes than extended households that are formed in the absence of the program. Clearly, in the absence of an impact of the program on living arrangement, one would not expect a change in the pool of children living with the elderly. A causal effect of the pension on living arrangements does not also imply that the pool of children living with the elderly is changed.

Why would selection explain part of the observed difference in health outcomes between children living with pensioners and children living with non-beneficiary elderly? Children co-residing with the elderly as a result of the program may differ in terms of health attributes from children who co-reside in the absence of the program. Suppose that the program increases the likelihood that an extended household be formed (as in Edmonds 2005) and that health outcomes of children living with pensioners are higher than health outcomes of children living with non-beneficiary elderly (as in Duflo 2003). Going back to the three mechanisms driving the effect of the program on living arrangement –preferences over household structure, change in family contractual arrangements, and efficiency gains — if these factors are correlated with child's health status, then sorting on child's health status is expected. For instance, co-residence of elderly and children may be more desirable for both the elderly and the children's parents when child's health is poor. Extended households that are formed as a result of the program then tend to foster less healthy children than extended household formed in the absence of the program.

4.4 Empirical Framework

The objective of the empirical analysis is to isolate the treatment effect from the selection effect of the pension on children's health outcomes. We start by defining the parameters of interest, and then lay out our estimation strategy, discussing the selection effect issue.

4.4.1 Parameters of interest

Our main outcome of interest, Y, is an anthropometric measure of child's health. The effect of the program on the measure of child's health may be direct, or through an effect on the child's living arrangement. Let $Z = 1$ if a child is living with an elderly relative, and 0 otherwise. Elderly relatives includes grandparents, great-grandparents, granduncle and grandaunt. We restrict the analysis to young children (those who have not yet reached their $6th$ anniversary) who have at least one of their four grandparent

alive and/or live with an elderly relative.[9]

In order to assess the effects of the program, we would ideally want to observe each child in the two states of the world, with program and without program. The central issue in evaluating the effect of this program is that we cannot observe $Y(0), Z(0)$ and $Y(1), Z(1)$, *i.e.*, the potential outcomes in the two states, for the same person. One person can only be observed in one of the two states. In the presence of the program, the only observed outcomes are $Y = Y(1)$ and $Z = Z(1)$, and $Y(0), Z(0)$ are counterfactual outcomes. Both potential outcomes are necessary to define the parameters of interest.

Since our analysis is at the child-level, we define treatment status at this level, even though the pension recipient is the elderly and not the child. Let $D = 1$ if the child has at least one grandparent who would be eligible to the pension program had the program been in place, and 0 otherwise. We distinguish between the following parameters:

- $\Delta_Y = E(Y_1 - Y_0|D = 1)$ is the causal effect of the pension program on all children who have at least one grand-parent eligible to the pension (irrespective of whether the child co-reside with him/her). The effect is not expected to be large in magnitude since nonresident grandparents who are eligible to the pension may not be contributing much to their grandchildren's health.
- $\Delta_Y|_{Z_1=1} = E(Y_1 - Y_0|D = 1, Z_1 = 1)$ is the causal effect of the pension program for those living with a pensioner.
- $\Delta_Z = E(Z_1 - Z_0|D = 1)$ is the causal effect of the program on children's living arrangements. This parameter is estimated in Edmonds *et al.* (2005). They find that eligible grandparents are more likely to be living with young children than non-eligible grandparents.

We now discuss identification and estimation of the first two parameters.

4.4.2 Estimation strategy

We first outline our identifying assumptions. Our estimation strategy exploits the fact that the age is the binding eligibility criteria for most of the Black South African households in our sample. We abstract from issues of selection into treatment by comparing children who have at least one grandparent who is age-eligible to the pension to children whose grandparents are not age-eligible. For that to help us identify the parameters of

[9] In the following, we will refer to relative elderly and grandparents as *grandparents*, for the sake of brevity.

interest, we assume that age randomly excludes potential participants from the pension program. We also assume that there are no take-up issues, no substitution bias and no general equilibrium effects. Let $T = 1$ if the child has at least one grandparent who is age-eligible to the pension. Then, the observed outcomes for $T = 0$ children can be used as a counterfactual for what $T = 1$ children would have experienced in the absence of the program. These standard assumptions are used in other work assessing the impact of the South African pension program (Case and Deaton, 1998; Duflo, 2000; Duflo, 2003; Edmonds et al., 2005).

Under these assumptions, we obtain:

$$\Delta_Y = E(Y_1 - Y_0|D = 1) = E(Y|D = 1, T = 1) - E(Y|D = 1, T = 0).$$

This parameter may be estimated as simple mean differences or within a linear regression framework. The equation of interest is:

$$Y = X'\beta + \Delta_Y T + \epsilon, \tag{4.1}$$

where X is a matrix of observable characteristics at the child and household levels, β the vector of parameters associated to X, and ϵ the error term. Since we have multiple observations by household, we allow for error terms to be correlated among children within the same households. The main parameter of interest is Δ_Y; in the result section, we present estimates for two specifications on X, with a small and large set of observables characteristics.

Identification of the causal effect of the pension program for those living with a pensioner, $\Delta_Y|_{Z_1=1}$, raises more challenges. We have the following result.

Lemma 1. *Let* $\Delta_S = E(Y_0|D = 1, Z_1 = 1) - E(Y_0|D = 1, Z_0 = 1)$. *Under the set of assumptions described above,* $\Delta_Y|_{Z_1=1} = E(Y|D = 1, Z = 1, T = 1) - E(Y|D = 1, Z = 1, T = 0)$ *if and only if* $\Delta_S = 0$.

Proof.

$$\begin{aligned}\Delta_Y|_{Z_1=1} &= \{E(Y_1|D = 1, Z_1 = 1) - E(Y_0|D = 1, Z_1 = 1)\} + E(Y_0|D = 1, Z_0 = 1) - E(Y_0|D = 1, Z_0 = 1) \\ &= \{E(Y_1|D = 1, Z_1 = 1) - E(Y_0|D = 1, Z_0 = 1)\} - \Delta_S.\end{aligned}$$

□

Let $\Delta_Y|_{Z=1} = \{E(Y_1|D = 1, Z_1 = 1) - E(Y_0|D = 1, Z_0 = 1)\}$. Under the assumption that the assignment of the program to the elderly is exogenous, we have:

$$\begin{aligned}\Delta_Y|_{Z=1} &= E(Y_1|D = 1, Z_1 = 1, T = 1) - E(Y_0|D = 1, Z_0 = 1, T = 0) \\ &= E(Y|D = 1, Z = 1, T = 1) - E(Y|D = 1, Z = 1, T = 0).\end{aligned}$$

Thus,

$$\Delta_Y|_{Z_1=1} = \{E(Y|D=1,Z=1,T=1) - E(Y|D=1,Z=1,T=0)\} - \Delta_S.$$

Hence, the result·

The standard exogeneity assumption on the assignment to the program is not sufficient to estimate the causal effect of the pension program for those living with a pensioner as the difference in *observed* outcomes for children living with eligible grandparents and those living with non-eligible grandparents. In other words, the observed outcome for children living with non-eligible grandparents is not a valid counterfactual to assess the situation that children who live with an eligible grandparent would have experienced in the absence of the pension program.

The problem comes from the fact that the program may have altered the *composition* of the pool of children residing with a grandparent. If, for instance, co-residence is more desirable when child's health is poorer, then the 'marginal' children who switch from no co-residence to coresidence with an elderly as a result of the program tend to be in better health than those who live with a grandparent in the absence of the program. In this case, the program makes it possible for children who are in better health to reside with a grandparent, and the selection effect is positive:

$$\Delta_S = E(Y_0|D=1,Z_1=1) - E(Y_0|D=1,Z_0=1) > 0.$$

Thus, unless there is no change in the composition of children living with eligible grandparents, $\Delta_Y|_{Z=1}$ provide a biased estimate of the causal effect of the pension program for those living with a pensioner, $\Delta_Y|_{Z_1=1}$. In our example, it would overestimate the effect of interest.

However, if a household had extended living arrangements prior to the birth of the child and remained extended after the birth of the child, then one may argue that the health status of the new born comes as a surprise to all household members. If there is no sorting into extended households based on the child's health status, the difference in average health outcomes of children born into households with eligible grandparents and those born into households with non-eligible grandparents then provides an unbiased estimate of the causal effect of the program for children born into households with a pensioner. When the decision to co-reside cannot be based on child's health, which is yet to be realized, we do not expect a change in the composition of a pool of children living with grandparents. In this case, the difference in average child health outcomes for children born into extended households with eligible and non-eligible grandparents may be interpreted as the causal effect of the pension program for those born into an extended household. If, in addition, this causal effect can be assumed to be the same for children born into extended households and all those co-residing with a pensioner, then, by restricting the sample to children born into extended household we can recover an estimate of the causal effect of interest, $\Delta_Y|_{Z_1=1}$.

To estimate the causal effect of the pension program for those born into an extended household, we use equation (4.1), limiting the sample to this set of children. To make our results comparable with previous results (Duflo, 2000, 2003), we estimate $\Delta_Y|_{Z=1}$ using equation (4.1), but restricting the sample only to children living with elderly, regardless of whether they were born into an extended household or not. Note, however, that we cannot use as Duflo (2000, 2003) the cohort of older children who were only partially exposed to the program as a comparison group. This limit to our estimation strategy is due to data constraints, as the majority of children for whom we observe anthropometric measures are born after the start of the program.

4.5 Data

This paper uses the Kwazalu Natal Income Dynamics Survey (KIDS).[10] Kwazalu-Natal is the largest province in South Africa (10 million inhabitants). Eighty percent of the population is Zulu. This survey builds on the Project for Statistics on Living Standards and Development (PSLSD) undertaken in 1993 and was designed to be representative at a provincial level. This population was re-surveyed in 1998 and 2004. To ensure comparability, the 1998 and 2004 questionnaires were designed to follow closely the 1993 version with some minor changes.

Questions were asked on household level, family and person level characteristics. The questionnaire covers a wide variety of topics: demography, household services, household expenditure, educational status and expenditure, remittances and marital maintenance, land and access use, employment and income, health status and expenditure and anthropometry. All children living in the selected households, were measured and weighted. Anthropometric measures were obtained for all children under 7 years in 1993 and under 14 years in 1998 that are present in the household. This paper uses two waves; 1993 and 1998.[11] We merge the first two waves (1993 and 1998) and construct a unique identifier for each individual.

In this paper, we focus the sample of households where children between the age of 6 months 5 years old (inclusive) are living. We are interested in extended living arrangements that include children, along with at least an adult from the older generation (grandparent, grand-aunt or grand-uncle). We are also interested in knowing whether the child was born in an extended living arrangement. We identify children who are co-residing with their grandparents (or great-grandparents) through an algorithm based on relationships to the head of households. We classified households as extended if the

[10]This panel is a collaborative project of the International Food Policy Research Institute, the University of Natal Durban, the University of Wisconsin-Madison and the South African Labor Development Research Unit at the University of Cape Town.

[11]During the period of the third wave period (i.e., 2004) Child Support Grants were provided to households with children. Identifying the effect of this program is beyond the scope of this paper.

family link is greater than or equal to 3.[12] As for nuclear household, they include at least one child and one adult parent and no older relative (parent, parent-in-law, aunt or uncle). Thus, using information identifying the head of household and the relationship to the head, we classified all households where young children are living (in wave 2) in one of these two categories .[13] Then given this information, we link the households to the first wave and assess the nature of the household structure at that time. By merging these two waves, we can flag households that became extended in the past five years and focus on children born to extended households. In contrast, (Duflo, 2000, 2003) focuses on children present in extended households, regardless of whether this household structure was already in place when the children were born. For each child, we determine whether he/she has a grandparent alive (as reported by the parents of the child), whether he/she is living with at least one grandparent or an elderly relative. We also usually observe the age of the elderly person. In the event the age of the elderly is not reported, we predict it as in Duflo (2003). [14]

Our objective is to assess the impact of the pension on children's health. This is why we focus on children anthropometric measures as dependent variables. We use height-for-age z-scores and weight-for-height z-scores. When the weight-for-height is 2 standard deviations below its expected value the child is wasted. Similarly when the height for age is 2 standard deviations below its expected value the child is stunted. In addition, Height-for-age is a long-term measure; it can serve as an ex-post indicator for inputs received in the first three years of life (Aguero et al., 2007). Consequently, insufficient height reflects a long-term malnutrition due to sustained unbalanced diets, poor health and protein deficiency. Weight-for-height, which is defined for each age (in months), measures the short-run nutritional status. Therefore, it allows to capture the effect of the recent presence of a pensioner in a household, an increase in calorie intake that is not yet translated in growth, or a recent increase in calorie intake. Given the South African context, we think that this measure is as important as height-for-age as it may capture an increase in calorie intake that is not necessarily due a protein balanced diet. If this is the case, then we may observe effects weight-for-height but not on height-for-age.

The descriptive statistics for height-for-age are provided in table 4.1 and descriptive statistics for weight-for-height table are provided in table 4.2.[15] According to the descriptive statistics, it appears that around 75% of children have an eligible grandmother and 35% of children have an eligible grandfather. When we focus on co-residence we see that 48% of children co-reside with an eligible female while only 17% of children

[12] We allow for children's parents to be absent from the household.

[13] After applying a crude algorithm that allowed us to classify the bulk of households, we checked the remaining by examining them on a case-by-case basis, exploiting the detailed information of the relationship to the household head.

[14] Following Duflo's work, a child is considered eligible if one of the following is true: the mother/father of the child is older than 34 years and her/his mother is still alive, or the mother/father is older than 32 years and her/his father is still alive.

[15] Since we do not have the exact sample for both we report their statistics separately.

co-reside with male elderly. This is not very surprising as women live longer and are more likely to co-reside with grand-children. As for the average number of school years of parents, it seems to be fairly similar between females and males: respectively 4.5 for females and 4.7 for males. On average children do not suffer from wasting or stunting the mean weight-for-height and height-for-age are respectively 0.662 and −0.745. In addition, the prevalence of wasting is around 22% and the prevalence of stunting is 3.5%.[16]

4.6 Results

4.6.1 Main Results

First we consider the impact of pension eligibility on children using exogenous controls regardless of living arrangements (i.e., Δ_Y). This parameter reflects the unconditional impact of pension eligibility on children health outcomes. Estimation results reported in table 4.3 show the coefficients for the impact of respectively female and male eligibility on children's weight-for-height z-scores (whz) and height-for-age z-scores (haz). The first three columns report results for weight-for-height z-scores and the last three columns for height-for-age z-scores. On a general level, pension eligibility seems to have no significant impact on children anthropometric measures. Having a female elderly eligible has a positive non significant impact on weight-for-height, and this impact remains insignificant and positive when we run separate regressions girls. Yet, it becomes negative for boys. As for its impact on height-for-age z-score, female eligibility has a positive and statistically insignificant impact that remains as such when separate regressions are computed by gender. Male pension eligibility seems to have a negative non-significant impact on weight-for-height and positive non-significant impact on height-for-age z-score. These impact remains in the same direction across gender though the impact on boys' height-for-age z-scores appear to be much higher than girls'.

The control variables reveal that age has a negative and statistically significant impact on height-for-age z-scores of -0.58 standard deviations and that these effects are not linear. This suggests that the older the child is the lower the height-for-age z-score. The height-for-age is a long term measure of previous nutrition deprivations. The older the child gets the more he will accumulate a stock of deprivation. More interestingly, these effects becomes higher for girls when we run separate regressions by gender. Father's age seem to have a negative effect on children in general and on girls in particular. A possible explanation for this impact is that older fathers might invest less in their daughters' health than young fathers do. Turning to the mother's age, it appears that it has a positive significant impact on girls' long-term health measure. The age of the mother can be considered as reflection of acquired experience in child

[16]Results are not reported here but can be provided upon request.

care and home allocation of resources. Thus, the older mothers might have better ways of allocating resources and taking care of young girls. Mother's education appears also to have a small positive significant impact on girls weight-for-height of 0.06 standard deviations.

The parameters estimated in table 4.3 must be interpreted with care for two main reasons. First, a child is considered as a beneficiary of the pension even if he/she is not co-residing with an age-eligible grandparent. Even if we assume that the non-resident grandparent lives close enough, there is no theoretical or empirical evidence that shows that one can treat co-residence as a special case of close proximity and pool both households together (Compton and Pollak, 2009). Besides, the non-resident grandparent may not even be living in close proximity and may not be contributing in any ways to his/her grandchild's health. Second, and as discussed in Section 3, there are potential confounding post treatment effects accruing from the impact of pension eligibility on co-residence itself (Edmonds et al., 2005). This leads us to focus our attention to children who are present in extended households.

The impact of pension eligibility in extended households (i.e., $\Delta_{Y|Z=1}$) is reported in table 4.4. Estimation results show that female eligibility have a positive significant effect of 0.51 on girls' weight-for-height and a negative non significant effect on boys. On average, co-residing with a female elderly increases weight-for-height by 0.50 standard deviation. The positive effect is in line with Duflo (2000, 2003)'s result. Yet the magnitude is different since we are using a different cohort from a different dataset (we only focus on one province) and at a different point in time. Turning our attention to the impact of male eligibility, we find that it has a positive – but non significant– impact on girls anthropometric measures (as in Duflo, 2003) and a positive significant effect on boy's height-for-age of 0.53. As for the correlates, they follow roughly the same pattern as the previous estimation. Mother's education and age are positively correlated with respectively weight-for-height and height-for-age. One explanation is that more educated mothers provide more food to young children, thus leading to an increase in weight-for-height and that older mothers have acquired more experience in child rearing and resource allocation within the household.[17] Once we run separate regressions for boys and girls we notice that the positive impacts of mother's education and age on weight-for-height z-score and height-for-age z-score hold only for girls. This suggests that as mothers have more education or grow older, they may develop a taste for discrimination in favour of girls. Turning our attention to father's age we notice that it is positively correlated with short term anthropometric measures.

These results, while they focus on a very specific definition of eligibility, must be interpreted with caution. Evidence in Edmonds et al. (2005) show that the pension has prompted adult children and their parents to form extended households. If an elderly's co-residence decision is correlated with children's health status, results from table 4.4 are most likely subject to selection bias. In this context, estimating a treatment effect

[17]It is also possible that older women may be giving birth to healthier children.

while conditioning on household structure may be misleading as we cannot disentangle the impact of the pension from the impact of the change in the household structure itself. As a result, the estimated parameters may overstate the impact of the pension and may be misleading.

Estimation results presented in table 4.5 account for this possibility. Specifically, we block the impact of the pension on household structure by focusing on children who are born in extended households.[18] We find that the impacts of the pension eligibility on children's anthropometric measures are still of the same sign but that the magnitudes changed and they are no longer significant. The impact of female eligibility on girls weight-for-height decreased by 23.5% (0.12 standard deviations) from 0.51 to 0.39 and lost its significance. Similarly, the impact of male eligibility on boys height-for-age z-score decreased by 34% from 0.53 to 0.35 standard deviations and became statistically insignificant.

To summarize, if we accept the hypothesis that the elderly's decision to co-reside prior to the birth of the child is independent from the child's health status then the results obtained while conditioning on household structure are biased upward. This result is consistent with a positive selection effect resulting from the fact that the program is making it possible for children of better health status to co-reside with a grandparent.

4.6.2 Further Results

The presence of significant impacts of pension eligibility depicted in extended households might be driven by the fairly modest number of covariates used. While the controls were selected for their purely exogenous nature it is important to see how the parameters of interest react when more controls (but not purely exogenous) are added. Table 4.7 shows estimation results with a richer set of controls. The impact of female eligibility and male eligibility on respectively girls weight-for-height and male height-for-age is still positive and statistically significant but smaller in magnitude. A similar regression was estimated to assess whether the absence of impacts depicted for children born to extended households was due to lack of controls. Estimation results presented in table 4.6 confirm that this is not the case.

While the impact of the pension eligibility on children health disappears when we block the impact of the treatment on the post-treatment variable, one can argue that the absence of pension's impact on children anthropometric measures may be due to a decrease in power as the sample size is reduced. To assess if this could possibly drive the absence of significant impacts of the pension, we run the same regression for children who were not born in extended household. The estimation results reported in table 4.8

[18] It should be noted that these results may not be representative.

show that even with a reduced sample size the impacts that we estimated earlier can be recovered. For children who were not born in extended households, but are living in an extended households, female eligibility has a positive significant effect on girls and male eligibility has a positive significant effect on boys.

4.7 Conclusion

In this paper we assess whether household composition may be a possible channel through which income may flow to affect children's health. To assess this possibility, we exploit the unique nature of a South African quasi-natural experiment (universal Old Age Pension) and use the dynamic nature of the data to identify changes in households' composition. We then use this information to block the impact of the pension on household composition and identify the net impact of the program on children anthropometric measures. If the change in household composition is an important channel through which income can produce its impacts on children health outcomes, then such cash transfers programs can be more effective if coupled with intervention that facilitate changes in living arrangements. Results suggest that the pension effect that was found previously was entirely due to changes in household composition. Such an effect is consistent with elderly pairing up with children of better health status under the program. This suggests that pension cash transfers targeted to the elderly poor do not always trickle down to children. It cautions against conditioning on variables that are themselves affected by a treatment when estimating an average treatment effect on beneficiaries.

There are many possible extensions to our work. The first natural extension would be to assess whether universal old age pensions in other developing countries has produced similar unintended effects on living arrangements and whether the impact of these pensions trickle down to children. Also, it would be interesting to assess the impact of the presence of grandparents on children's anthropometric measures in developing countries where there are no universal pension programs. The impact of family dissolution on children's health would be another interesting avenue to explore. Finally, one could augment all these extensions by assessing the long-term on children's school performance.

Table 4.1: Summary statistics for the variable used in regressions for height-for-age z-scores

Variable	Mean	S.D
Eligible Female	0.733	0.443
Eligible Male	0.353	0.478
Co-resident Eligible Female	0.479	0.5
Co-resident Eligible Male	0.175	0.38
Age	3.162	1.554
Age Squared	12.409	10.246
Gender	0.495	0.5
Household size (Co-residents)	10.569	5.011
Household size (All)	12.257	5.625
Number of children ages < 6	2.507	1.498
Sex ratio	1.491	1.075
Father's Age	38.582	6.242
Mother's Age	29.284	6.542
Mother's Education	4.478	3.253
Father's Education	4.742	4.385
Number of children ages 6-14	3.033	2.062
Number of children ages 15-24	2.805	1.998
Num. of members ages 25-49	3.739	2.037
Females ages 50-54 present	0.175	0.38
Females ages 55-59 present	0.125	0.331
Males ages 50-54 present	0.125	0.331
Males ages 55-59 present	0.1	0.3
Males ages 60-64 present	0.069	0.253
Mean haz children 6-14	-0.395	0.573
haz	-0.745	2.125
Obs		641

Table 4.2: Summary statistics for variables used in regression for weight-for-height z-scores

Variable	Mean	S.D
Eligible Female	0.736	0.441
Eligible Male	0.35	0.477
Co-resident Eligible Female	0.482	0.5
Co-resident Eligible Male	0.172	0.377
Age	3.16	1.554
Age squared	12.398	10.225
Gender	0.494	0.5
Household size (Co-residents)	10.56	5.007
Households size (All)	12.278	5.634
Number of children < 6	2.514	1.501
Sex ratio	1.494	1.08
Father's Age	38.609	6.227
Mother's Age	29.213	6.519
Mother's Education	4.478	3.253
Father's Education	4.742	4.385
Number of children ages 6-14	3.022	2.049
Number of children ages 15-24	2.819	1.998
Number of children ages 25-49	3.733	2.038
Females ages 50-54	0.17	0.376
Females ages 55-59	0.127	0.333
Males ages 50-54	0.122	0.328
Males ages 55-59	0.102	0.303
Males ages 60-64	0.067	0.25
Mean whz children 6-14	0.195	0.541
whz	0.662	1.708
Obs		629

Table 4.3: The Impact Pension eligibility on Children Anthropometric Measures for all Households Structures

	All whz	Girls whz	Boys whz	All haz	Girls haz	Boys haz
Female Eligible	0.0932	0.2572	-0.0435	0.2559	0.0977	0.3070
	(0.1808)	(0.2123)	(0.2586)	(0.2101)	(0.2822)	(0.2622)
Male Eligible	-0.2230	-0.2395	-0.1710	0.2820	0.1494	0.3578
	(0.1875)	(0.2428)	(0.2442)	(0.1961)	(0.2880)	(0.2706)
Age	-0.0697	0.1697	-0.2655	-0.5812**	-0.9484**	-0.2634
	(0.2204)	(0.3051)	(0.3132)	(0.2434)	(0.3659)	(0.3119)
Age Squared	0.0034	-0.0296	0.0309	0.0737**	0.1203**	0.0350
	(0.0329)	(0.0463)	(0.0473)	(0.0346)	(0.0531)	(0.0439)
Gender	-0.0378			-0.2170		
	(0.1281)			(0.1598)		
Father's Age	0.0244*	0.0318	0.0196	-0.0282*	-0.0413*	-0.0215
	(0.0141)	(0.0246)	(0.0157)	(0.0162)	(0.0214)	(0.0220)
Mother's Age	0.0196	0.0037	0.0307	0.0445**	0.0850***	0.0177
	(0.0159)	(0.0220)	(0.0218)	(0.0181)	(0.0248)	(0.0263)
Mother's Education	0.0384	0.0626*	0.0218	0.0335	0.0060	0.0389
	(0.0265)	(0.0332)	(0.0409)	(0.0273)	(0.0421)	(0.0364)
Father's Education	0.0009	-0.0058	0.0161	0.0322	0.0074	0.0548
	(0.0252)	(0.0349)	(0.0367)	(0.0334)	(0.0474)	(0.0397)
Cons	-0.6965	-1.0412	-0.5166	-0.7852	-0.4013	-1.2712
	(0.7043)	(1.0310)	(0.8171)	(0.8090)	(1.0766)	(1.0538)
Obs.	629	318	311	641	324	317
R-Squared	0.0266	0.0340	0.0328	0.0419	0.0837	0.0286

Table 4.4: The impact of Pension Eligibility on Anthropometric Measures in Extended Households

	All whz	Girls whz	Boys whz	All haz	Girls haz	Boys haz
Female Eligible	0.0625	0.5073**	-0.3348	0.1830	0.0574	0.2527
	(0.1724)	(0.2236)	(0.2201)	(0.1886)	(0.2660)	(0.2461)
Male Eligible	-0.0548	-0.1665	0.0476	0.3502	0.1713	0.5254*
	(0.2211)	(0.2973)	(0.2607)	(0.2473)	(0.4014)	(0.2724)
Age	-0.1717	0.1757	-0.3485	-0.4939**	-0.7042*	-0.3628
	(0.2367)	(0.3360)	(0.3246)	(0.2496)	(0.3879)	(0.3097)
Age Squared	0.0208	-0.0289	0.0495	0.0568	0.0825	0.0435
	(0.0355)	(0.0512)	(0.0491)	(0.0356)	(0.0557)	(0.0441)
Gender	-0.0663			-0.2065		
	(0.1375)			(0.1703)		
Father's Age	0.0304*	0.0457	0.0258	-0.0073	-0.0136	-0.0081
	(0.0167)	(0.0308)	(0.0163)	(0.0183)	(0.0237)	(0.0247)
Mother's Age	0.0171	0.0025	0.0240	0.0448**	0.0778***	0.0213
	(0.0171)	(0.0232)	(0.0225)	(0.0192)	(0.0252)	(0.0280)
Mother's Education	0.0425	0.0735**	0.0221	0.0315	0.0048	0.0447
	(0.0275)	(0.0348)	(0.0416)	(0.0284)	(0.0445)	(0.0369)
Father's Education	0.0058	0.0119	0.0082	0.0616	0.0549	0.0620
	(0.0279)	(0.0440)	(0.0401)	(0.0391)	(0.0655)	(0.0425)
Cons	-0.8144	-1.8124	-0.3865	-1.6068	-1.6530	-1.5410
	(0.8110)	(1.2931)	(0.8448)	(0.9126)	(1.3383)	(1.1270)
Obs	563	279	284	572	283	289
R-Squared	0.0278	0.0566	0.0402	0.0448	0.0806	0.0356

Table 4.5: The Impact Pension Eligibility on Anthropometric Measures of Children Born to Extended Households

	All whz	Girls whz	Boys whz	All haz	Girls haz	Boys haz
Female Eligible	0.0061	0.3882	-0.3661	0.2351	0.3450	0.0909
	(0.2038)	(0.2599)	(0.2815)	(0.2205)	(0.2966)	(0.3190)
Male Eligible	0.0598	-0.0880	0.1821	0.2576	0.1429	0.3514
	(0.2433)	(0.3332)	(0.2905)	(0.2403)	(0.4240)	(0.2546)
Age	-0.1181	0.3298	-0.4223	-0.4658	-0.6768	-0.2777
	(0.3043)	(0.3823)	(0.4613)	(0.3228)	(0.4745)	(0.4078)
Age Squared	0.0151	-0.0438	0.0552	0.0512	0.0720	0.0357
	(0.0446)	(0.0587)	(0.0672)	(0.0453)	(0.0683)	(0.0565)
Gender	-0.1123			-0.1821		
	(0.1609)			(0.1923)		
Father's Age	0.0260	0.0352	0.0239	-0.0263	-0.0120	-0.0379
	(0.0204)	(0.0361)	(0.0206)	(0.0215)	(0.0273)	(0.0270)
Mother's Age	0.0175	0.0094	0.0181	0.0561**	0.0871***	0.0329
	(0.0202)	(0.0260)	(0.0268)	(0.0215)	(0.0287)	(0.0315)
Mother's Education	0.0372	0.0711*	0.0166	0.0294	0.0030	0.0312
	(0.0333)	(0.0410)	(0.0533)	(0.0336)	(0.0507)	(0.0436)
Father's Education	0.0105	0.0194	0.0102	0.0319	0.0238	0.0384
	(0.0347)	(0.0544)	(0.0520)	(0.0380)	(0.0636)	(0.0474)
Cons	-0.7599	-2.0133	0.0828	-1.2713	-2.1193	-0.7077
	(0.9913)	(1.5259)	(1.2073)	(1.1528)	(1.6572)	(1.3884)
Obs	426	215	211	433	218	215
R-Squared	0.0252	0.0474	0.0405	0.0510	0.0965	0.0352

Table 4.6: The Impact Pension eligibility on Children Anthropometric Measures Born in Extended Households Structures

	All whz	Girls whz	Boys whz	All haz	Girls haz	Boys haz
Eligible Female	0.1189	0.3144	-0.0110	0.0223	0.1458	0.0313
	(0.2275)	(0.2926)	(0.3497)	(0.2573)	(0.3496)	(0.4042)
Eligible Male	-0.0494	-0.0771	-0.1075	0.1684	-0.1003	0.3537
	(0.2563)	(0.3677)	(0.3115)	(0.2355)	(0.4094)	(0.2818)
Age	-0.1732	0.2184	-0.4548	-0.4256	-0.6025	-0.1841
	(0.2926)	(0.3742)	(0.4510)	(0.3412)	(0.4972)	(0.4145)
Age Squared	0.0257	-0.0282	0.0646	0.0450	0.0592	0.0218
	(0.0427)	(0.0561)	(0.0659)	(0.0485)	(0.0714)	(0.0586)
Gender	-0.1928			-0.3012		
	(0.1712)			(0.1973)		
Mean whz Sibling	0.8416***	0.8096***	0.8507***			
	(0.1864)	(0.2629)	(0.2872)			
Mean haz Sibling				0.6291***	0.3740	0.7754***
				(0.2191)	(0.3005)	(0.2807)
Sex Ratio	-0.0752	-0.2050	0.2119	-0.0700	-0.0558	-0.1316
	(0.0890)	(0.0972)	(0.2476)	(0.1213)	(0.1448)	(0.2362)
Number of children< 5	-0.0584	-0.0799	-0.0286	-0.1300	-0.3296**	0.0146
	(0.0839)	(0.1274)	(0.1045)	(0.1066)	(0.1424)	(0.1467)
HH Size Residents	-0.0407	-0.0581	-0.0419	0.0676	0.1621**	0.0205
	(0.0397)	(0.0607)	(0.0496)	(0.0501)	(0.0687)	(0.0647)
Father's Age	0.0284	0.0343	0.0197	-0.0329	-0.0222*	-0.0533
	(0.0193)	(0.0302)	(0.0240)	(0.0236)	(0.0295)	(0.0286)
Mother's Age	0.0083	0.0057	0.0074	0.0597***	0.0904	0.0446
	(0.0197)	(0.0236)	(0.0276)	(0.0220)	(0.0302)	(0.0313)
Mother's Education	0.0257	0.0541	0.0155	0.0296	0.0235	-0.0086
	(0.0321)	(0.0420)	(0.0535)	(0.0349)	(0.0511)	(0.0488)
Father's Education	-0.0002	0.0136	-0.0224	0.0417	0.0121	0.0675
	(0.0306)	(0.0499)	(0.0495)	(0.0362)	(0.0527)	(0.0489)
Num of children 6-14	-0.0020	-0.0546	0.0381	0.0074	0.0047	-0.0264
	(0.0533)	(0.0833)	(0.0738)	(0.0747)	(0.0989)	(0.0952)
Num of children 15-24	0.0373	0.0075	0.0351	0.0402	0.0459	0.0728
	(0.0574)	(0.0908)	(0.0645)	(0.0624)	(0.1013)	(0.0765)
Hous. Member 25-49	0.0627	0.1040	0.0384	-0.0450	-0.1343	-0.0194
	(0.0642)	(0.0682)	(0.0994)	(0.0680)	(0.0841)	(0.0984)
Women Age 50-54	0.2455	0.3255	0.3340	0.1754	0.3959	-0.0816
	(0.3013)	(0.4253)	(0.3903)	(0.3334)	(0.4784)	(0.4173)
Women Age 55-59	-0.1682	-0.5769**	0.3189	-0.4345	-0.3193	-0.3797
	(0.2209)	(0.2759)	(0.3594)	(0.2916)	(0.3967)	(0.4687)
Men Age 50-54	0.3232	0.1524	0.4419	-0.4899	-1.3159**	0.7473
	(0.3017)	(0.4042)	(0.5295)	(0.4700)	(0.6099)	(0.6659)
Men age 55-59	0.0820	0.1521	0.0139	-0.0724	-0.3773	0.0715
	(0.3279)	(0.3700)	(0.5719)	(0.3127)	(0.4441)	(0.4121)
Men Age 60-64	0.2147	0.6726	-0.1117	-0.4565	-1.0483	-0.0456
	(0.4131)	(0.8162)	(0.3832)	(0.3970)	(0.5450)	(0.6507)
Cons	-0.2613	-0.8151	0.2725	-0.8983	-1.7201*	-0.3965
	(0.9599)	(1.4925)	(1.2609)	(1.1877)	(1.7346)	(1.2788)
Obs	426	215	211	433	218	215
R-Squared	0.1229	0.1776	0.1410	0.1063	0.1946	0.1273

Table 4.7: The Impact Pension eligibility on Children Anthropometric Measures Present in Extended Households Structures

	All whz	Girls whz	Boys whz	All haz	Girls haz	Boys haz
Eligible Female	0.0962	0.4134*	-0.2264	0.1499	0.0149	0.3478
	(0.2121)	(0.2474)	(0.2920)	(0.2088)	(0.2804)	(0.3147)
Eligible Male	-0.1658	-0.2669	-0.1624	0.3384	0.2051	0.4778*
	(0.2276)	(0.3146)	(0.2685)	(0.2476)	(0.3924)	(0.2849)
Age	-0.1655	0.1168	-0.2983	-0.5575**	-0.8451**	-0.3575
	(0.2238)	(0.3046)	(0.3160)	(0.2672)	(0.4075)	(0.3203)
Age Squared	0.0213	-0.0220	0.0454	0.0664*	0.1017*	0.0435
	(0.0334)	(0.0457)	(0.0480)	(0.0385)	(0.0592)	(0.0465)
Gender	-0.0882			-0.2629		
	(0.1452)			(0.1775)		
Mean whz Sibling	0.8504***	0.7878***	0.8846***			
	(0.1555)	(0.1960)	(0.2584)			
Mean Haz Sibling				0.6635***	0.4171	0.8662**
				(0.2076)	(0.2651)	(0.2698)
Sex Ratio	-0.0127	-0.0583	0.0202	-0.0282	-0.0159	-0.0296
	(0.0636)	(0.0816)	(0.1715)	(0.0834)	(0.1020)	(0.1402)
Number of children < 5	-0.0070	-0.0135	0.0292	-0.1777*	-0.2640**	-0.1305
	(0.0795)	(0.1113)	(0.0961)	(0.0969)	(0.1339)	(0.1367)
HH Size Residents	-0.0354	-0.0375	-0.0525	0.0686	0.0824	0.0815
	(0.0361)	(0.0546)	(0.0437)	(0.0476)	(0.0665)	(0.0653)
Father's Age	0.0304*	0.0430	0.0267	-0.0114	-0.0135	-0.0181
	(0.0159)	(0.0266)	(0.0185)	(0.0187)	(0.0239)	(0.0250)
Mother's Age	0.0110	0.0032	0.0158	0.0532**	0.0823***	0.0368
	(0.0171)	(0.0219)	(0.0236)	(0.0193)	(0.0259)	(0.0273)
Mother's Education	0.0341	0.0591	0.0202	0.0309	0.0144	0.0216
	(0.0279)	(0.0365)	(0.0443)	(0.0289)	(0.0446)	(0.0416)
Father's Education	-0.0055	0.0032	-0.0151	0.0804**	0.0763	0.0902**
	(0.0258)	(0.0401)	(0.0405)	(0.0373)	(0.0654)	(0.0448)
Num of children 6-14	-0.0140	-0.0456	0.0340	0.0713	0.1258	0.0031
	(0.0505)	(0.0700)	(0.0640)	(0.0704)	(0.0975)	(0.0867)
Num of children 15-24	0.0167	-0.0390	0.0556	0.0797	0.1275	0.0329
	(0.0499)	(0.0782)	(0.0598)	(0.0614)	(0.0899)	(0.0735)
Hous. Member 25-49	0.0562	0.0726	0.0618	-0.0740	-0.1133	-0.0731
	(0.0542)	(0.0602)	(0.0781)	(0.0596)	(0.0795)	(0.0826)
Women Age 50-54	0.1703	0.2018	0.1337	0.1210	0.4272	-0.1269
	(0.2316)	(0.3310)	(0.3071)	(0.2726)	(0.3769)	(0.3506)
Women Age 55-59	-0.1986	-0.5315**	0.1518	-0.1183	-0.1797	0.0127
	(0.1886)	(0.2437)	(0.2809)	(0.2776)	(0.3229)	(0.4483)
Men Age 50-54	0.1359	0.0411	0.1952	0.0509	-0.5621	0.6815
	(0.2220)	(0.3053)	(0.3406)	(0.3410)	(0.4964)	(0.4643)
Men Age 55-59	-0.0086	-0.0910	0.0071	-0.1018	-0.1149	-0.1963
	(0.2669)	(0.2981)	(0.4279)	(0.2724)	(0.3696)	(0.3501)
Men Age 60-64	0.0594	0.3013	-0.3081	-0.4308	-0.7971*	-0.0432
	(0.3613)	(0.6221)	(0.3591)	(0.3704)	(0.4659)	(0.6795)
Cons	-0.5054	-1.0711	-0.3841	-1.8294**	-1.9046	-1.8609*
	(0.8208)	(1.2284)	(0.8972)	(0.9078)	(1.4123)	(1.0073)
Obs	563	279	284	572	283	289
R-Squared	0.1138	0.1590	0.1231	0.1002	0.1592	0.1181

Table 4.8: The Impact Pension Eligibility on Anthropometric Measures of Children not Born to Extended Households

	All whz	Girls whz	Boys whz	All haz	Girls haz	Boys haz
Female Eligible	0.5824	1.6376*	-0.2327	0.1145	-1.2331	1.0066**
	(0.5279)	(0.8506)	(0.5150)	(0.5298)	(0.8023)	(0.5089)
Male Eligible	-0.7059	-0.5045	-0.9491	1.2668	0.3831	1.8723*
	(0.6260)	(0.5821)	(0.7473)	(0.9872)	(1.0161)	(1.1170)
Age	-0.1611	-0.0529	-0.3033	-0.8183**	-1.3068**	-0.5231
	(0.3553)	(0.7040)	(0.4817)	(0.4002)	(0.5468)	(0.5273)
Age Squared	0.0065	-0.0220	0.0687	0.1303**	0.2076**	0.0449
	(0.0602)	(0.1122)	(0.0887)	(0.0659)	(0.0838)	(0.0860)
Gender	0.0504			-0.2769		
	(0.2897)			(0.3617)		
Father's Age	0.0349	0.0841	0.0174	0.0500*	-0.0170	0.0660*
	(0.0277)	(0.0673)	(0.0272)	(0.0268)	(0.0508)	(0.0276)
Mother's Age	0.0431	-0.0292	0.0642*	-0.0328	0.0474	-0.0471
	(0.0341)	(0.0446)	(0.0354)	(0.0376)	(0.0470)	(0.0486)
Mother's Education	0.0561	0.0804	0.0596	0.0546	-0.0154	0.0737
	(0.0401)	(0.0781)	(0.0565)	(0.0492)	(0.0720)	(0.0707)
Father's Education	0.0008	-0.0019	0.0175	0.1439	0.1759	0.0993
	(0.0452)	(0.0783)	(0.0701)	(0.0839)	(0.1314)	(0.0836)
Cons	-1.8329	-1.6074	-2.0460	-0.9726	0.2671	-1.6274
	(1.6181)	(3.3339)	(1.4718)	(1.4263)	(2.6716)	(1.7892)
Obs	137	64	73	139	65	74
R-Squared	0.0965	0.2577	0.1047	0.1262	0.2264	0.1940

Chapter 5

Conclusion

This thesis provides an investigation of the mechanisms through which peer effects in social networks (friends and family) flow. A higher understanding of these pathways provides insights that allows policy makers to exploit social multiplier adequately and provide an improved and efficient policy intervention. While pursuing this objective I focused on three questions:

1. Are peer effects in obesity mediated by the presence of peer effects in eating habits?
2. Are peer effects in health outcomes mediated by the presence of peer effects in physical activity?
3. In the presence of a change in a social network, can the estimated average treatment effect on the treated be mediated while conditioning on a post treatment variable?

The first question can be viewed as the cornerstone of my interest in the transmission mechanisms of peer effects in adolescents' behaviour. While the presence of peer effects in obesity provides a valuable information, it does not give grounds for non-discriminatory counter-obesity policy intervention. However, if one can identify a behavioural channel (e.g., eating habits) through which these effects flow, then one can use the social multiplier to reduce the spread of obesity. In the second chapter, I investigate whether the presence of peer effects in adolescent weight gain is mediated by the presence of peer effects in fast food consumption. I first study the presence of endogenous peer effects in fast food consumption. New methods based on spatial econometric analysis are used to identify and estimate this model, under the assumption that individuals interact through a friendship social network. Our results indicate that an increase in his friends' mean fast food consumption induces an adolescent to

increase his own fast food consumption. This peer effect amplifies through a social multiplier the impact of any exogenous shock on fast food consumption. Our estimated social multiplier is 1.59. In addition to assessing the presence of peer effects in fast food consumption, I relate fast food consumption on adolescent's body mass index through a dynamic production function. The results indicate that fast food consumption has a positive impact on weight gain. Specifically, a one-unit increase in the weekly frequency (in days) of fast food consumption produces an increase in BMI by 1.5% within a year. This effect reaches 2.4% when the social multiplier is taken into account. All this is consistent with the hypothesis that if there are peer effects in obesity a large part of them flow through eating habits. Therefore, the presence of a social multiplier coupled with reduction of the relative price of fast food and increasing availability of fast food restaurants over time could exacerbate the prevalence of obesity in the years to come.

I address the second question in the third chapter where I focus on the transmission mechanism through which peer effects in health outcomes can flow. I first assess the presence of endogenous peer effects in physical activity. The methods I use in this chapter are inspired from the ones I used in the previous chapter. Results from the peer effect model are consistent with the presence of a social multiplier in physical activity of 1.12. A friend's mean exercising frequency increases an adolescent's exercising frequency. This suggests that an individual's lifestyle is affected by his peer's lifestyle. To assess the impact of physical activity on health outcome, I estimate a health production function. The results show that exercising has a positive significant impact on health of 0.09 (i.e., 4 %). This impact increases to 0.10 (i.e., 5 %) in the short run and 0.13 (i.e., 6%) in the long-run once I account for the social multiplier. With the increase in the availability of equipments that encourage sedentary lifestyles (e.g., video games, social networking on computers), the presence of the social multiplier might exacerbate the impact of an exogenous shock on adolescents' exercising habits and consequently their health as adults. This in turn, increases the incidence of chronic disease imposing high health care costs as well as economic costs (decrease in productivity).

One might be left with the impression that the social multiplier can only be attributed to "bad" outcomes yet, this is not true. If adequately identified, the social multiplier can contribute to the decline of the spread of the epidemics overviewed earlier and decrease health care costs. This will be the case as long as it is exploited by policy makers through tax and subsidy reforms encouraging healthy habits among adolescents.

Last but not least, I answer the third question in chapter four where social network is defined as household composition rather than friends. This definition is a little different from the one I used in the previous chapters but still conveys the same idea: the importance of the role of transmission mechanisms in social networks. Living arrangements are variable to a certain extent (e.g., marriage, extended living arrangements) yet, in the presence of financial incentives, additional changes might occur. For instance, an exogenous shock on one's income may produce an impact on a person who is indirectly

related to him, or through a change in the composition of the network. In this case, the shock may produce its effects directly or indirectly or even through a combination of both events. I assess the extent to which such indirect effects are likely to occur in the context of a developing country: South Africa. I exploit the unique nature of their old age pension quasi-natural experiment and its well established impacts on children's heath and household composition. If the change in household composition is an important channel through which income can produce its impacts on children health outcomes then such programs can be more effective if coupled with intervention that facilitate changes in living arrangements. Results suggest that the pension effect that was found previously was entirely due to changes in household composition. Such an effect is consistent with elderly pairing up with children of better health status under the program. This in turn suggests that pension cash transfers impacts are not due to a trickle down to children effect but rather selection bias. It also cautions against conditioning on post-treatment variables that are themselves affected by the assessed treatment.

-Extensions

There are many possible extensions to the work that was done in this thesis. One possible extension would be to allow for endogenous peer effects to vary by gender. Consequently, it would be useful to develop a general approach that would allow same sex and opposite sex peer effects to be different for both males and females. Also, it would be interesting to take the two step link formation approach a step further. Therefore, finding instrumental variables that directly affect the formation of social networks but does not affect the outcomes of interest would be valued. To my knowledge no work has been done in this direction yet.

Bibliography

Abu-Abid, S., Szold, A., and Klausner, J. (2002). Obesity and cancer. *Journal of medicine*, 33(1-4):73.

Adams, P., Hurd, M., McFadden, D., Merrill, A., and Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, 112(1):3–56.

Adler, N., Boyce, T., Chesney, M., Cohen, S., Folkman, S., and Kahn, R. (1994). Socioeconomic Status and Health. The Challenge of the Gradient. *American Psychologist*, 49(1):15–24.

Aguero, J. M., Carter, M. R., and Woolard, I. (2007). The impact of unconditional cash transfers on nutrition: The south african child support grant. Working Papers 39, International Poverty Center.

Aguiar, M. and Hurst, E. (2005). Consumption versus expenditure. *Journal of Political Economy*, 113(5):919–948.

Akerlof, G. (1991). Procrastination and obedience. *The American Economic Review*, pages 1–19.

Andersen, R., Crespo, C., Bartlett, S., Cheskin, L., and Pratt, M. (1998). Relationship of physical activity and television watching with body weight and level of fatness among children. *JAMA: The Journal of the American Medical Association*, 279(12):938.

Anderson, K. (2000). Family structure, parental investment, and educational outcomes among black South Africans. *PSC Research Report 00-461. Population Studies Center, University of Michigan.*

Anderson, M. and Matsa, D. (2011). Are restaurants really supersizing america? *American Economic Journal: Applied Economics*, 3(1):152–188.

Anderson, P., Butcher, K., and Levine, P. (2003). Maternal employment, child care and childhood obesity. *Journal of Health Economics*, 22(3):477–505.

Angrist, B., Bettinger, E., and Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review*, 96(3):847–862.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *The American Economic Review*, 92(5):1535–1558.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297.

Auld, M. and Powell, L. (2008). Economics of Food Energy Density and Adolescent Body Weight. *Economica*, 15:1468–0335.

Babcock, P. and Hartman, J. (2010). Networks and Workouts: Treatment Size and Status Specific Peer Effects in a Randomized Field Experiment. Technical report, National Bureau of Economic Research.

Benjamins, M., Hummer, R., Eberstein, I., and Nam, C. (2004). Self-reported health and adult mortality risk: an analysis of cause-specific mortality. *Social science & medicine*, 59(6):1297–1306.

Berentzen, T., Petersen, L., Schnohr, P., and Sørensen, T. (2008). Physical activity in leisure-time is not associated with 10-year changes in waist circumference. *Scandinavian Journal of Medicine & Science in Sports*, 18(6):719–727.

Berlin, J. and Colditz, G. (1990). A meta-analysis of physical activity in the prevention of coronary heart disease. *American journal of epidemiology*, 132(4):612.

Blanchflower, D. and Oswald, A. (2009). Imitative obesity and relative utility. *Journal of the European Economic Association*.

Bleich, S., Cutler, D., Murray, C., and Adams, A. (2007). Why is the developed world obese? Working Paper 12954, National Bureau of Economic Research.

Bleich, S., Cutler, D., Murray, C., and Adams, A. (2008). Why is the developed world obese? *Annual Review of Public health*, 29:273–295.

Block, J., Scribner, R., and DeSalvo, K. (2004). Fast food, race/ethnicity, and income A geographic analysis. *American Journal of Preventive Medicine*, 27(3):53–67.

Blume, L., Brock, W., Durlauf, S., and Ioannides, Y. (2010). *Identification of Social Interactions*. North Holland.

Bowman, S., Gortmaker, S., Ebbeling, C., Pereira, M., and Ludwig, D. (2004). Effects of fast-food consumption on energy intake and diet quality among children in a national household survey. *Pediatrics*, 113(1):112–118.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.

Brock, W. A. and Durlauf, S. N. (2001). Interactions-based models. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5 of *Handbook of Econometrics*, chapter 54, pages 3297–3380. Elsevier.

Burman, S. (1995). Child care by the elderly and the duty of support in multigenerational households. *Southern African Journal of Gerontology*, 4(1):13–17.

Burman, S. (1996). Intergenerational family care: legacy of the past, implications for the future. *Journal of Southern African Studies*, 22(4):585–598.

Calabr, P., Golia, E., Maddaloni, V., Malvezzi, M., Casillo, B., Marotta, C., Calabrò, R., and Golino, P. (2009). Adipose tissue-mediated inflammation: the missing link between obesity and cardiovascular disease? *Internal and Emergency Medicine*, 4(1):25–34.

Calle, E. (2007). Obesity and cancer. *British Medical Journal*, 335(7630):1107–1108.

Caraher, M. and Cowburn, G. (2007). Taxing food: implications for public health nutrition. *Public Health Nutrition*, 8(08):1242–1249.

Carrell, S., Fullerton, R., and James, E. (2008). ÒDoes Your Cohort Matter? Measuring Peer Effects in College Achievement,Ó. *NBER Working Paper*, 14032.

Carrell, S., Hoekstra, M., and West, J. (2011). Is poor fitness contagious? Evidence from randomly assigned friends. *Journal of Public Economics*.

Case, A. (2001). Does money protect health status? Evidence from South African pensions.

Case, A. and Deaton, A. (1998). Large cash transfers to the elderly in south africa. *Economic Journal*, 108(450):1330–61.

Case, A., Lubotsky, D., and Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, pages 1308–1334.

Chandrasekhar, A. and Lewis, R. (2011). Econometrics of sampled networks. *MIT working paper*.

Charness, G. and Gneezy, U. (2009). Incentives to Exercise. *Econometrica*, 77(3):909–931.

Chen, S. E., Florax, R. J., and Snyder, S. D. (2009). Obesity in urban food markets: Evidence from geo-referenced micro data. 2009 Annual Meeting, July 26-28, 2009, Milwaukee, Wisconsin 49512, Agricultural and Applied Economics Association.

Chou, S.-Y., Rashad, I., and Grossman, M. (2005). Fast-food restaurant advertising on television and its influence on childhood obesity. NBER Working Papers 11879, National Bureau of Economic Research.

Christakis, N. and Fowler, J. (2007a). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379.

Christakis, N. and Fowler, J. (2007b). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379.

Christakis, N. and Fowler, J. (2010). Examining dynamic social networks and human behavior. *Preprint.*

Christakis, N., Fowler, J., Imbens, G., and Kalyanaraman, K. (2010). An empirical model for strategic network formation. Technical report, National Bureau of Economic Research.

Clark, A. and Loheac, Y. (2007). It Wasnt Me, It Was Them! *Journal of Health Economics*, 26:763–784.

Cliff, A. and Ord, J. (1981). *Spatial processes: models & applications.* Pion Ltd.

Cohen, S. (2004). Social relationships and health. *American Psychologist*, 59(8):676.

Cohen-Cole, E. and Fletcher, J. (2008a). Are all health outcomes contagious? detecting implausible social network effects in acne, height, and headaches. *British Medical Journal*, 3337:a2533.

Cohen-Cole, E. and Fletcher, J. (2008b). Is obesity contagious?: social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics*, 27(5):1143–1406.

Compton, J. and Pollak, R. (2009). Proximity and coresidence of adult children and their parents: description and correlates. *Mimeo- University of Michigan Retirement Research Center.*

Contoyannis, P., Jones, A., and Rice, N. (2004). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics*, 19(4):473–503.

Courtemanche, C. (2007). A Silver Lining? The Connection Between Gasoline Prices and Obesity. *Economic Inquiry.*

Currie, A., Shields, M. A., and Price, S. W. (2007). The child health/family income gradient: Evidence from england. *Journal of Health Economics*, 26(2):213–232.

Currie, J., DellaVigna, S., Moretti, E., and Pathania, V. (2010). The effect of fast food restaurants on obesity and weight gain. *American Economic Journal: Economic Policy*, 2:34–65.

Currie, J. and Stabile, M. (2003). Socioeconomic status and child health: Why is the relationship stronger for older children? *American Economic Review*, pages 1813–1823.

Cutler, D., Glaeser, E., and Sphapiro, J. (2003). Why have americans become more obese? *Journal of Economic Perspectives*, 17:93–118.

De Agostini, P., of Essex. Institute for Social, U., and Research, E. (2007). *Diet composition, socio-economic status and food outlets development in Britain.* Institute for Social and Economic Research.

Deaton, A. (2002). Policy implications of the gradient of health and wealth.

Deaton, A. and Paxton, C. (1998). Aging and inequality in health and income. *American Economic Review*, 88(2):248–253.

Dinardo, J. Garlick, R. and Strange, . K. (2010). The elusive causal effec of body weight on mortality and health. *Mimeo.*

Duflo, E. (2000). Child health and household resources in south africa: Evidence from the old age pension program. *American Economic Review*, 90(2):393–398.

Duflo, E. (2003). Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in south africa. *World Bank Economic Review*, 17(1):1–25.

Dunn, R. A. (2008). Obesity and the Availability of Fast-Food: An Instrumental Variables Approach. *SSRN eLibrary.*

Durlauf, S. N. (2004). *Neighborhood effects*, volume 4 of *Handbook of Regional and Urban Economics.* Elsevier.

Edmonds, E., Mammen, K., and Miller, D. (2005). Rearranging the family? household composition responses to large pension receipts. *Journal of Human Resources*, 40(1):186–207.

Fiszbein, A., Schady, N., and Ferreira, F. (2009). *Conditional cash transfers: reducing present and future poverty.* World Bank Publications.

Fletcher, J. (2010). Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics*, 19(4):466–484.

Fletcher, J. and Ross, S. (2011). Estimating the effects of friendship networks on health behaviors of adolescents. *Working Papers.*

Flores, C. and Flores-Lagunes, A. (2007). Identification and estimation of causal mechanisms and net effects of a treatment. *Mimeo-University of Florida.*

Folsom, A., Kushi, L., and Hong, C. (2000). Physical activity and incident diabetes mellitus in postmenopausal women. *American Journal of Public Health*, 90(1):134.

Ford, E., Williamson, D., and Liu, S. (1997). Weight change and diabetes incidence: findings from a national cohort of US adults. *American Journal of Epidemiology*, 146(3):214.

Fortin, B. and Yazbeck, M. (2011). Peer Effects, Fast Food Consumption and Adolescent Weight Gain. *Cahier de recherche/Working Paper*, 11:03.

Fowler, J. and Christakis, N. (2008a). The dynamic spread of happiness in a large social network. *BMJ (Clinical research ed.)*, 337:a2338.

Fowler, J. and Christakis, N. (2008b). Estimating peer effects on health in social networks. *Journal of Health Economics*, 27(5):1400–1405.

Fowler, J. and Christakis, N. (2008c). Estimating peer effects on health in social networks: A response to cohen-cole and fletcher; trogdon, nonnemaker, pais. *Journal of health economics*, 27(5):1400.

Frijters, P., Haisken-DeNew, J., and Shields, M. (2005). The causal effect of income on health: Evidence from German reunification. *Journal of Health Economics*, 24(5):997–1017.

Gaviria, A. and Raphael, S. (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, 83(2):257–268.

Gillman, M., Rifas-Shiman, S., Berkey, C., Field, A., and Colditz, G. (2003). Maternal gestational diabetes, birth weight, and adolescent obesity. *Pediatrics*, 111(3):e221.

Glaeser, E., Sacerdote, B., and Scheinkman, J. (2003). The social multiplier. *Journal of the European Economic Association*, 1(2-3):345–353.

Hamer, M. and Chida, Y. (2008). Walking and primary prevention: a meta-analysis of prospective cohort studies. *British journal of sports medicine*, 42(4):238.

Haydon, A., MacInnis, R., English, D., and Giles, G. (2006). Effect of physical activity and body size on survival after diagnosis with colorectal cancer. *Gut*, 55(1):62.

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900.

Helmrich, S., Ragland, D., Leung, R., and Paffenbarger Jr, R. (1991). Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *New England Journal of Medicine*, 325(3):147–152.

Hsiao, C. (1981). Autoregressive modelling and money-income causality detection. *Journal of Monetary Economics*, 7(1):85–106.

Hu, F., Manson, J., Stampfer, M., Colditz, G., Liu, S., Solomon, C., and Willett, W. (2001). Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *New England Journal of Medicine*, 345(11):790–797.

Hu, F., Sigal, R., Rich-Edwards, J., Colditz, G., Solomon, C., Willett, W., Speizer, F., and Manson, J. (1999). Walking compared with vigorous physical activity and risk of type 2 diabetes in women. *JAMA: the journal of the American Medical Association*, 282(15):1433.

Hu, G., Lindstrom, J., Valle, T., Eriksson, J., Jousilahti, P., Silventoinen, K., Qiao, Q., and Tuomilehto, J. (2004). Physical activity, body mass index, and risk of type 2 diabetes in patients with normal or impaired glucose regulation. *Archives of internal medicine*, 164(8):892.

Hu, G., Qiao, Q., Silventoinen, K., Eriksson, J., Jousilahti, P., Lindstr "om, J., Valle, T., Nissinen, A., and Tuomilehto, J. (2003). Occupational, commuting, and leisure-time physical activity in relation to risk for type 2 diabetes in middle-aged finnish men and women. *Diabetologia*, 46(3):322–329.

Idler, E. and Kasl, S. (1991). Health perceptions and survival: Do global evaluations of health status really predict mortality? *Journal of gerontology*, 46(2):S55.

Jaccard, J., Blanton, H., and Dodge, T. (2005). Peer influences on risk behavior: an analysis of the effects of a close friend. *Developmental Psychology*, 41(1):135–147.

Jacobson, M. and Brownell, K. (2000). Small taxes on soft drinks and snack foods to promote health. *American Journal of Public Health*, 90(6):854–857.

Jago, R., Macdonald-Wallis, K., Thompson, J., Page, A., Brockman, R., and Fox, K. (2010). Better With a Buddy: The Influence of Best Friends on Children's Physical Activity. *Medicine and Science in Sports and Exercise*.

Jeffery, R. and French, S. (1998). Epidemic obesity in the United States: are fast foods and television viewing contributing? *American Journal of Public Health*, 88(2):277.

Jekanowski, M., Binkley, J., and Eales, J. (2001). Convenience, accessibility, and the demand for fast food. *Journal ofAgricultural and Resource Economics*, 26(1):58–74.

Kelejian, H. and Prucha, I. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121.

Kelejian, H. and Prucha, I. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, pages 509–533.

Kelejian, H. and Prucha, I. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157:53–67.

Khanam, R., Nghiem, H., and Connelly, L. (2009). Child health and the income gradient: Evidence from Australia. *Journal of Health Economics*, 28(4):805–817.

Kim, D. and Kawachi, I. (2006). Food taxation and pricing strategies to thin out the obesity epidemic. *American journal of preventive medicine*, 30(5):430–437.

Kohl III, H. and Hobbs, K. (1998). Development of physical activity behaviors among children and adolescents. *Pediatrics*, 101(3):549.

Kuh, D. and Ben-Shlomo, Y. (2004). *A life course approach to chronic disease epidemiology*, volume 2. Oxford University Press, USA.

Lakdawalla, D. and Philipson, T. (2002). The growth of obesity and technological change: A theoretical and empirical examination. Working Paper 8946, National Bureau of Economic Research.

Lee, L. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22(4):307–335.

Lee, L.-F., Liu, X., and Lin, X. (2010). Specification and estimation of social interaction models with networks structure. *The Econometrics Journal*, 13(2):143–176.

Lee, Y. (2000). The predictive value of self assessed general, physical, and mental health on functional decline and mortality in older adults. *Journal of Epidemiology and Community Health*, 54(2):123.

Levitsky, D., Halbmaier, C., and Mrdjenovic, G. (2004). The freshman weight gain: a model for the study of the epidemic of obesity. *International Journal of Obesity*, 28(11):1435–1442.

Li, C., Engstrom, G., Hedblad, B., Calling, S., Berglund, G., and Janzon, L. (2006a). Sex differences in the relationships between BMI, WHR and incidence of cardiovascular disease: a population-based cohort study. *International Journal of Obesity*, 30(12):1775–1781.

Li, T., Rana, J., Manson, J., Willett, W., Stampfer, M., Colditz, G., Rexrode, K., and Hu, F. (2006b). Obesity as compared with physical activity in predicting risk of coronary heart disease in women. *Circulation*, 113(4):499.

Lin, B., Guthrie, J., and Blaylock, J. (1996). *The Diets of America's Children: Influence of Dining Out, Household Characteristics, and Nutrition Knowledge*. US Dept. of Agriculture, Economic Research Service.

Lin, X. (2010). Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. *Journal of Labor Economics*, 28(4):825–860.

Lindahl, M. (2005). Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. *Journal of Human Resources*, 40(1):144.

Lyons, R. (2011). The spread of evidence-poor medicine via flawed social-network analysis. *Statistics, Politics, and Policy*, 2(1):2.

Maggio, C. and Pi-Sunyer, F. (2003). Obesity and type 2 diabetes. *Endocrinology and metabolism clinics of North America*, 32(4):805–822.

Magnus, K., Matroos, A., and Strackee, J. (1979). Walking, cycling, or gardening, with or without seasonal interruption, in relation to acute coronary events. *American Journal of Epidemiology*, 110(6):724.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–42.

Marmot, M. and Wilkinson, R. (2006). *Social determinants of health.*

Marshall, T. (2000). Exploring a fiscal food policy: the case of diet and ischaemic heart disease. *BMJ: British Medical Journal*, 320(7230):301.

McCracken, V. and Brandt, J. (1987). Household consumption of food-away-from-home: total expenditure and by type of food facility. *American Journal of Agricultural Economics*, 69(2):274–284.

McGee, D., Liao, Y., Cao, G., and Cooper, R. (1999). Self-reported health status and mortality in a multiethnic us cohort. *American Journal of Epidemiology*, 149(1):41.

McInnes, M. M. and Shinogle, J. A. (2009). Physical activity: Economic and policy factors. Working Paper 15039, National Bureau of Economic Research.

McLellan, L., Rissel, C., Donnelly, N., and Bauman, A. (1999). Health behaviour and the school environment in New South Wales, Australia. *Social Science & Medicine*, 49(5):611–619.

Meer, J., Miller, D., and Rosen, H. (2003). Exploring the health–wealth nexus. *Journal of health Economics*, 22(5):713–730.

Meyerhardt, J. et al. (2006a). Impact of physical activity on cancer recurrence and survival in patients with stage III colon cancer: findings from CALGB 89803. *Journal of clinical oncology*, 24(22):3535.

Meyerhardt, J. et al. (2006b). Physical activity and survival after colorectal cancer diagnosis. *Journal of clinical oncology*, 24(22):3527.

Moffit, R. (2001). *Policy Interventions, Low equilibria, and Social Interaction (Social Dynamics)*. MIT press.

Moller, V. and Ferreira, M. (2003). Getting by: Benefits of non-contributory pension income for older south african households. *Mimeo.*

Morris, J., Clayton, D., Everitt, M., Semmence, A., and Burgess, E. (1990). Exercise in leisure time: coronary attack and death rates. *British Heart Journal*, 63(6):325.

Niemeier, H., Raynor, H., Lloyd-Richardson, E., Rogers, M., and Wing, R. (2006). Fast food consumption and breakfast skipping: predictors of weight gain from adolescence to adulthood in a nationally representative sample. *Journal of Adolescent Health*, 39(6):842–849.

Nocon, M., Hiemann, T., M
"uller-Riemenschneider, F., Thalau, F., Roll, S., and Willich, S. (2008). Association of physical activity with all-cause and cardiovascular mortality: a systematic review and meta-analysis. *Journal of Cardiovascular Risk*, 15(3):239.

Noel, H. and Nyhan, B. (2011). The "unfriending" problem: The consequences of homophily in friendship retention for causal estimates of social influence. *Social Networks*, 33(3), pages = (211 - 218).

Okada, K., Hayashi, T., Tsumura, K., Suematsu, C., Endo, G., and Fujii, S. (2000). Leisure-time physical activity at weekends and the risk of type 2 diabetes mellitus in japanese men: the osaka health survey. *Diabetic Medicine*, 17(1):53–58.

Oswald, A. and Powdthavee, N. (2007). Obesity, unhappiness, and the challenge of affluence: Theory and evidence. *Economic Journal*, 117(521):441–454.

Paffenbarger, R., WING, A., and HYDE, R. (1978). Physical activity as an index of heart attack risk in college alumni. *American Journal of epidemiology*, 108(3):161.

Pisinger, C., Toft, U., Aadahl, M., Gl
"umer, C., and Jørgensen, T. (2009). The relationship between lifestyle and self-reported health in a general population:: The inter99 study. *Preventive medicine*, 49(5):418–423.

Pollak, R. (1988). Tied transfers and paternalistic preferences. *The American Economic Review*, 78(2):240–244.

Powell, K., Thompson, P., Caspersen, C., and Kendrick, J. (1987). Physical activity and the incidence of coronary heart disease. *Annual review of public health*, 8(1):253–287.

Powell, L., Auld, M., Chaloupka, F., OMalley, P., and Johnston, L. (2007). Access to fast food and food prices: relationship with fruit and vegetable consumption and overweight among adolescents. *Adv Health Econ Health Serv Res*, 17:23–48.

Powell, L. and Bao, Y. (2009). Food Prices, Access to Food Outlets and Child Weight Outcomes: A Longitudinal Analysis. *Economics and Human Biology*, 7:64–72.

Powell, L., Tauras, J., and Ross, H. (2005). The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior. *Journal of Health Economics*, 24(5):950–968.

Propper, C., Rigg, J., and Burgess, S. (2007). Child health: evidence on the roles of family income and maternal mental health from a uk birth cohort. *Health Economics*, 16(11):1245–1269.

Renna, F., Grafova, I. B., and Thakur, N. (2008). The effect of friends on adolescent body weight. *Economics and Human Biology*, 6(3):377–387.

Robinson, T., Borzekowski, D., Matheson, D., and Kraemer, H. (2007). Effects of fast food branding on young children's taste preferences. *Archives of Pediatrics and Adolescent Medicine*, 161(8):792.

Rogers, P., Fusinski, K., Rathod, M., Loiler, S., Pasarica, M., Shaw, M., Kilroy, G., Sutton, G., McAllister, E., Mashtalir, N., et al. (2007). Human adenovirus Ad-36 induces adipogenesis via its E4 orf-1 gene. *International Journal of Obesity*, 32(3):397–406.

Rolfes, S. and Whitney, E. (1996). Understanding nutrition. *St. Paul: West Publishing Company*.

Rosenheck, R. (2008). Fast food consumption and increased caloric intake: a systematic review of a trajectory towards weight gain and obesity risk. *Obesity Reviews*, 9(6):535–547.

Ruhm, C. (2000). Are Recessions Good for Your Health?*. *Quarterly Journal of Economics*, 115(2):617–650.

Sallis, J. (2000). Age-related decline in physical activity: a synthesis of human and animal studies. *Medicine & Science in Sports & Exercise*, 32(9):1598.

Salvy, S.-J., Vartanian, L. R., Coelho, J. S., Jarrin, D., and Pliner, P. P. (2008). The role of familiarity on modeling of eating and food consumption in children. *Appetite*, 50(2-3):514 – 518.

Samad, A., Taylor, R., Marshall, T., and Chapman, M. (2005). A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Disease*, 7(3):204–213.

Shalizi, C. and Thomas, A. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2):211.

Sofi, F., Capalbo, A., Cesari, F., Abbate, R., and Gensini, G. (2008). Physical activity during leisure time and primary prevention of coronary heart disease: an updated meta-analysis of cohort studies. *Journal of Cardiovascular Risk*, 15(3):247.

Sorensen, H., Sabroe, S., Rothman, K., Gillman, M., Fischer, P., and Sorensen, T. (1997). Relation between weight and length at birth and body mass index in young adulthood: cohort study. *Bmj*, 315(7116):1137.

Steeg, G. V. and Galstyan, A. (2011). Ruling out latent homophily in social networks. *Mimeo*.

Stinebrickner, R. and Stinebrickner, T. (2006a). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics*, 90(8-9):1435–1454.

Stinebrickner, R. and Stinebrickner, T. R. (2006b). What can be learned about peer effects using college roommates? evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics*, 90:1435–1454.

Story, M., Neumark-Sztainer, D., and French, S. (2002). Individual and environmental influences on adolescent eating behaviors. *Journal of the American Dietetic Association*, 102(3S):40–51.

Trogdon, J. G., Nonnemaker, J., and Pais, J. (2008). Peer effects in adolescent overweight. *Journal of Health Economics*, 27(5):1388–1399.

Van Doorslaer, E. et al. (1997). Income-related inequalities in health: some international comparisons. *Journal of Health Economics*, 16(1):93–112.

VanderWeele, T. (2011). Sensitivity analysis for contagion effects in social networks. *Sociological Methods & Research*, 40(2):240.

Warburton, D., Nicol, C., and Bredin, S. (2006). Health benefits of physical activity: the evidence. *Canadian Medical Association Journal*, 174(6):801.

Wolin, K., Yan, Y., Colditz, G., and Lee, I. (2009). Physical activity and colon cancer prevention: a meta-analysis. *British journal of cancer*, 100(4):611–616.

Yakusheva, O., Kapinos, K., and Weiss, M. (2010). Peer effects and the Freshman 15: Evidence from a natural experiment. *Economics & Human Biology*.

Yang, X., Telama, R., Leino, M., and Viikari, J. (1999). Factors explaining the physical activity of young adults: the importance of early socialization. *Scandinavian journal of medicine & science in sports*, 9(2):120–127.

Zheng, H., Orsini, N., Amin, J., Wolk, A., Nguyen, V., and Ehrlich, F. (2009). Quantifying the dose-response of walking in reducing coronary heart disease risk: meta-analysis. *European journal of epidemiology*, 24(4):181–192.