The ambiguous causal relationship between body-mass and labour income in emerging economies: The case of Mexico.

Levasseur, Pierre

October 2017

The ambiguous causal relationship between body-mass and labour income in emerging economies: The case of Mexico.

Pierre Levasseur
Université de Bordeaux, GREThA
October, 2017
Preliminary version, please contact the author before quoting or citing
pierre.levasseur@u-bordeaux.fr

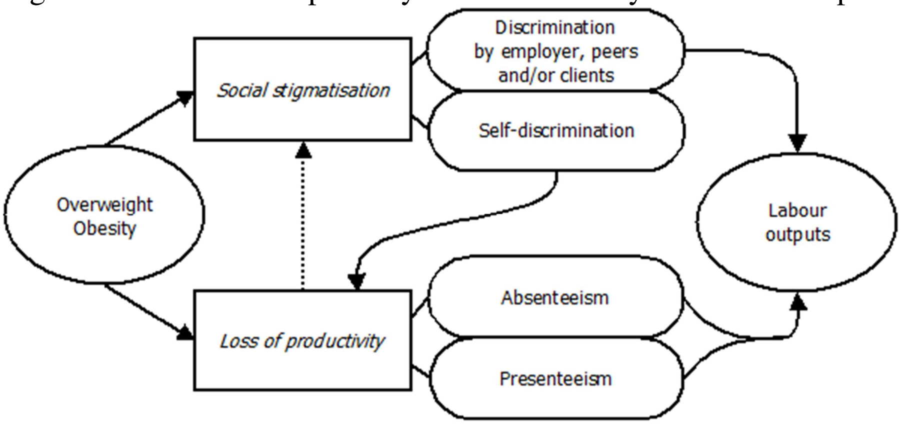
***Abstract**: The effect of body-mass on labour outcomes seems to be closely linked to the level of development of the country concerned. In rich countries, excess weight is penalised at work, whereas in the poorest societies overweight is rewarded. These divergences indicate that this effect depends on sociocultural factors related to weight perception and stigmatisation. In the case of emerging economies, weight perception appears to be unclear given a hybrid nutritional panorama: hunger and obesity coexist. Although the literature suggests a quadratic causal relationship between body-mass and earnings in middle-income countries, these economies are quite heterogeneous in terms of the body-mass distribution. The main objective of this study is therefore to explore the impact of body-mass index (BMI) on hourly income in an emerging country with high obesity prevalence, such as Mexico. We use panel data from the Mexican Family Life Survey and perform a bootstrapped three-step parametric model, based on an expanded Mincer earning function to control for potential sample selection bias and endogeneity problems. Then, we test the robustness of results implementing a bootstrapped three-step semiparametric model. For employees, our results show a right-leaning U-inverted causal relationship between BMI and hourly wage in Mexico. In other words, while overweight is rewarded at work, obesity is significantly penalised. By contrast, for self-employed workers, we observe a linear and positive effect of BMI on earnings, at least, up to a BMI of 32 kg/m². The source of stigmatisation, as well as sociocultural heterogeneity, can explain why earning penalties are higher for employers than self-employed workers. To conclude, our findings suggest that two paradoxical phenomena are occurring in emerging countries such as Mexico: (i) a social acceptance of overweight, due to past nutritional deprivations and the growing normalisation of obesity; (ii) a social reject of obesity, due to the large diffusion and adoption of thinness ideals from Western culture.*

***Keywords:** Mexico; emerging countries; labour income; obesity; weight stigma.*

1. INTRODUCTION

Since the number of individuals who are overweight and obese is increasing constantly around the world and reaching endemic levels in some countries (Egypt, Mexico, United States, South Africa, etc.), a growing number of researchers are becoming interested in the impact of these nutritional changes on labour outputs. As Cawley (2004) notes: bodyweight may affect employment and earnings through two channels: *social stigmatisation* and *loss of productivity*. Both pathways are schematised in Figure 1. The *social stigmatisation* process has various repercussions on labour opportunities. While the employer (or the clients) can discriminate an obese worker, choosing not to employ him, not to promote him and/or not to pay him fairly for his level of competence (or not to buy his product/service), the labour environment (ties with peers) forms a social space where an obese worker might be excluded (e.g. exclusion of project collaboration, poor networking). In addition, stigmatised individuals may suffer from psycho-sociological disorders, such as lack of self-esteem and self-confidence. Cawley (2004) uses the term of self-discrimination to describe this type of psycho-sociological disorders because it is the individual himself who restrains in labour opportunities. Furthermore, overweight and obesity are related to various non-transmissible diseases (diabetes, heart coronary attacks, cancers, etc.), as well as chronic fatigue syndrome (sleeping apnoea, inefficient brain oxygenation). Both bad physical and mental health of obese workers might directly drive on a *loss of productivity*, through absenteeism at work and presenteeism (loss of productivity during the working day).[1]

Figure 1: Transmission pathways between obesity and labour outputs



Source: Author.

In theory, the impact of the *loss of productivity* on professional success can only be negative. However, the process of *social stigmatisation* seems to have ambiguous effects on employment and earnings. The existing literature suggests that sociocultural beliefs related to

[1] We do not exclude the possibility that the *loss of productivity* drives on *social stigmatisation*, hence the dotted arrow in Figure 1. Indeed, it is likely that physical and socio-psychological inabilities of obese people do not meet expectations of employers, peers and/or clients and lead to a social reject from them.

weight perception are particularly important insofar as the effect of body-mass on the professional success depends on the level of a country's development. For instance, several authors note that the causal relationship between bodyweight and earnings is generally negative in rich economies, especially for women, despite differences between ethnicity and work status (Cawley 2004; Sarlio-Lahteenkorva, Silventoinen and Lahelma 2004; Brunello and D'Hombres 2007; Johar and Katayama 2012). These findings can be explained by the Western perception of weight. In these countries, obese people are generally perceived as lazy, slow, unattractive and lacking in rigour (Runge, 2007). By contrast, in the poorest societies where the hunger issue is still important, fat people are usually considered the most attractive because high body-mass symbolises health, strength, wealth, and prosperity (Renzaho, 2004). Therefore, being overweight or obese is not particularly associated with income penalties and, in some activities and sociocultural contexts, bodyweight tends to improve earnings and employment status, especially for men (Poulain, 2017). This is, for example, the case in sub-Saharan African countries such as Guinea, Ivory Coast, Ghana and Ethiopia (Glick and Sahn 1998; Schultz 2003; Kedir 2008).

In the case of emerging economies, the relationship does not appear to be so clear-cut. Given the coexistence of hunger and obesity in these countries, Shimokawa (2008) assumes a quadratic causal relationship between body-mass and labour outcomes. According to his results, the hypothesis appears to be valid in the Chinese context: underweight and obese workers are the lowest paid. More recently, analysing the impact of body-mass on employment probability in China, Pan, Qin and Liu (2013) relate the same quadratic relationship: normal-weight adults are more likely to work than their thinner and fatter counterparts.[2]

[2] In other emerging countries, results are closely related to the kind of employment. For instance, Dinda et al. (2006) show that being overweight is associated with higher wages for Indian coalminers, whereas findings from the Philippines suggest that overweight self-employed women earn significantly less than their slimmer counterparts (Colchero and Bishai 2012).

Although the literature suggests a quadratic causal relationship between body-mass and professional success in middle-income economies, this group of countries is quite heterogeneous in terms of nutritional issues. For example, obesity prevalence is three times higher in Mexico than in China. Consequently, it is likely that different nutritional contexts lead to different weight perception and, by extension, to different effects of body-mass on labour outputs. It thus seems difficult to speculate on the nature of this effect in countries such as Mexico, where hunger has virtually disappeared and being overweight has become the norm. Our main objective is therefore to determine the effect of body-mass index (BMI) on labour opportunities in Mexico, and more particularly on earnings. The analysis uses panel data from the *Mexican Family Life Survey* (MxFLS), which provides three survey waves (2002, 2006 and 2012), and focuses on the working-age population, especially on the earnings of employees and self-employed workers. Based on an expanded Mincer earning function, we implement a bootstrapped three-step parametric model (i.e. selection, instrumental variables and structural equations) in order to control for potential sample selection bias and endogeneity problems. We use the median BMI in the municipality to instrument the individual BMI. Both linear and quadratic specifications are analysed. Then, we run a bootstrapped three-step semiparametric model to perform a visual check of the causal relationship, without being limited by the traditional linear and quadratic hypotheses concerning its functional form. Our findings suggest that the causal relationship between body-mass and hourly wage takes a right-leaning U-inverted form in Mexico. In other words, the relationship is positive up to overweight status and becomes negative from the obesity threshold. Interestingly, this holds true only for employees. For self-employed workers, we observe a linear and positive effect of BMI on earnings, at least up to a BMI of 32 kg/m^2.

The rest of the paper is organised as follows: Section 2 describes the original conceptual framework used in this study, Section 3 establishes the methods, Section 4 presents the results, and Section 5 concludes and discusses the main findings.

2. CONCEPTUAL FRAMEWORK

In line with Cawley (2004), we consider two fundamental pathways through which obesity may affect labour outcomes such as wages: *loss of productivity* and *social stigmatisation* (Figure 1). As pointed out by Poulain (2017), this dual approach makes it possible to distinguish medical factors that are space and time-invariant (diseases, physical incapacities) from social factors that are subject to perpetual change (social stigmatisation). The local sociocultural specificities, denoted *S*, are supposed to determine how weight is perceived and stigmatised in a given society at a given time. Therefore, we adopt the following conceptual framework, where the effect of obesity on labour outcome is the sum of the effects of the *loss of productivity* and the *social stigmatisation* of weight (function *F* is assumed linear).

$$Labour\ Outcomes_S = F(productivity\ \ ; stigmatization_S)$$

The effect of the productivity pathway on professional success does not depend on sociocultural aspects (*S*) and is always supposed to be negative. It casts no doubt that the loss of productivity due to obesity-related health problems leads to worse labour outputs. However, the effect of weight stigmatisation on labour outcomes is ambiguous, since it depends on the system of values in place (*S*). In societies where obesity is a major health problem and hunger has completely disappeared, such as in Western countries, excess weight appears to be discriminating and penalising at work. Reversely, in sociocultural environments where hunger remains prevailing, as in the poorest societies, excess weight is well perceived and thus rewarded. In other words, concerning the most vulnerable populations, the social preference for overweight more than compensates for the loss of productivity associated with this nutritional status.

Nonetheless, the role of weight discrimination in employment and labour income is unclear in a sociocultural environment (S) where obesity is constantly rising and hunger persists, such as in emerging countries. In this particular context, we speculate three possible scenarios. First, cultural factors due to past nutritional deprivations may still influence weight perception. As in the poorest economies, the weight would be viewed as a sign of health and strength and be rewarded (Brewis, 2003). Second, in view of the rise of media and the spread of Western culture since the 1980s, a progressive adoption of the thinness ideals is occurring in developing countries (Brewis et al., 2011). Thus, the positive appreciation of obesity may tend to disappear given the convergence of the beauty standards around the world. In other words, the causal relationship between body-mass and labour outcomes would gradually pass from positive to negative during this process of sociocultural homogenisation. This scenario can explain why some authors observe a quadratic relationship in China, the inflection point of this curve being around normal-weight status (Shimokawa, 2008; Pan, Qin and Liu, 2013). This U-inverted curve would symbolize the transition from a system of values where overweight is preferred on the labour market to a situation where thinness is sought and leads to professional success. Finally, a third scenario reveals a contemporary (or future) change in weight perception (and stigmatisation) since the recent generalization of overweight and obesity in several rich and emerging countries (Classen, 2017). For example in the United States, Robinson and Christiansen (2014) observe a process of social acceptance of obesity in areas where overweight has become the physical norm. It is likely that this process is also occurring in emerging countries such as Mexico, where overweight and obesity rates reach the same level as the United States. In this case, the social environment would progressively become more tolerant concerning overweight given the increasing number of individuals with this nutritional status.

In fact, we expect that these three scenarios are concomitant in emerging economies. For instance, it can be assumed that overweight is socially accepted (scenarios 1 and 3), but only up to a certain threshold of body-mass. Indeed, an excessive weight may increase the risk of social stigmatisation, due to the progressive integration of Western stereotypes of beauty (scenario 2). If this hypothesis is verified, the causal relationship between body-mass and hourly income would take a right-leaning U-inverted form: the turning point being between overweight and obesity status (instead of normal-weight status as it is the case in China).

3. METHODS

3.1. Data and sample

The data used in this study come from the *Mexican Family Life Survey* (MxFLS), the first survey with a representative sample of the Mexican population at national, rural-urban and regional levels. Sampling directives were drawn up by the Mexican Institute of Statistics. The survey covers a 10-year period with three distinct waves. The first wave was carried out in 2002 with 35,677 individuals surveyed in 8,440 households (living in 150 municipalities across 16 Mexican states). Given the longitudinal dimension of the survey, the second (2005-06) and the third (2009-12) waves are based on the initial sample from 2002. For both waves, the follow-up rate of the initial sample exceeds 90% at the household level. MxFLS data include detailed information on socioeconomic characteristics of households and individuals. Moreover, anthropometric data were collected for all household members directly in the home by trained staff from the Mexican Institute of Public Health. Weight was measured with a set of digital scales (accuracy of 0.1 kg) and height with a stadiometer (accuracy of 0.5 cm).

Our sample is restricted to comply with the objective of the study. First, pregnant and lactating women were withdrawn from the sample in order to limit anthropometric bias. Second, in line with the labour economics literature, we analyse only the Mexican working-age population from 15 to 65 years old (Gong, Soest and Villagomez, 2004). Third, we mainly focus on the hourly earnings gap between workers. The advantage of this labour output is to capture aspects that others indicators (e.g. employment probability) cannot identify, as, for example, the presenteeism pathway (see Figure 1).[3] Therefore, the study is restricted to employees and self-employed workers, which represents 44.5% and 12.5% of the working-age population in Mexico, respectively (Table A.1 of the Appendix).[4] Employers are excluded because of the few observations available in this category (only 3% of the working-age population). Self-employed workers and employees are analysed individually as the former group receives profits and the latter receives salaries. Moreover, in theory, the sources of discrimination differ between both kinds of occupation. While obese employees are generally stigmatised by their employer and/or colleagues, self-employed workers can only be discriminated by clients. We should note that all obese workers (employees, self-employed and employers), in particular women, are potentially sensitive to what Cawley (2004) calls the self-discrimination, linked to a lack of confidence and self-esteem.

2.2. An expanded Mincer earning function

Insofar as the study is based on earning determinants, the general framework of the model takes an expanded form of the Mincer earning function, as follows (Mincer, 1974):

$$log\, Y_{it} = F(X_{it},\, B_{it}) \qquad (1)$$

Where Y_{it} is the hourly earnings of an individual *i* at the period *t*. In line with Vogl (2014), we

[3] We also implement additional estimations of the effect of body-mass on employment probability using an *IV-Probit* estimator in order to treat endogenous relationship between both factors. Results from these estimations will be presented in the part 3.3.

[4] The MxFLS reports five work statuses in Mexican cities: (i) inactive and unemployed people; (ii) unpaid workers; (iii) employees; (iv) self-employed workers; (v) employers.

measure the hourly earnings variable using the income from the principal activity during a regular month and the number of working hours during a regular week from this principal activity.[5] Taking into account the number of working hours allows the absenteeism pathway to be controlled for. In other words, the effect of body-mass on hourly earnings can only pass through the social stigmatisation pathway and/or the presenteeism pathway (see Figure 1). The hourly earnings are expressed in Mexican pesos using the year 2002 and the Centre-South region as the baseline in order to neutralize inflation and price gaps between regions. Like Campos-Vazquez, López-Calva, and Lustig (2016), we drop hourly incomes below and above the 1st and the 99th percentile from the sample to exclude extreme values. Finally, the hourly earnings' variable is log-transformed. Hence, only individuals with a log-hourly income between 0.6 and 5.1 are analysed (i.e. between 2 and 166 Mexican pesos).

Then, X_{it} refers to control variables that structure the earning function (function *F(.)* is not yet known). Inspired by the approach of Mulatu and Schooler (2002) and Nordman and Roubaud (2009), several individual determinants of labour income are included as control variables: age, the square of the age (proxy of experience), gender, marital status (in a couple or not), years of schooling, cognitive skills' score (measured using the mental test described by Raven 2000). We also incorporate in the earning function a variable counting the number of children in the household to identify how much the worker has to deal with budgetary constraints. Then, as suggested by Baum and Ford (2004), we add variables that control for survey years and area differences: a score of infrastructural development of the municipality given by the highest authority available in the municipality (i.e. the president or the vice-president of the municipality),[6] the region (South-East or not) and the geographic area (rural area, small city, middle-sized city or large city). Although it could be relevant to incorporate more information

---

[5] Hourly wage $= \frac{\text{monthly wage}}{\frac{\text{weekly working hours}}{7} \times 30{,}5}$.

[6] The score varies from 0 for a low level to 6 for a high level of infrastructural development, taking into account the presence (or absence) of public transportation, health centre, refuse collection, sewage system, hydraulic system and hard roads in the municipality.

concerning the quality of employment and the occupation into the earning function, this practice is extremely controversial in literature and can turn out counterproductive. Because occupational variables are strongly correlated with income and body-mass, the inclusion of such variables is likely to introduce endogeneity into the model and thus bias the estimates. Hence, under the guidance of Nordman and Roubaud (2009), we do not include this type of variables in the model.[7]

Finally, a body-mass indicator (B_{it}) is added to the set of regressors in order to extend the earning function to health and nutritional topics. The individual body-mass is measured using the body-mass index (BMI=weight(kg)/[height(m)]²). According to the World Health Organization (WHO 2000), an individual is underweight when his BMI is lower than 18.5 kg/m^2, normal-weight when his BMI is between 18.5 and 25 kg/m^2, overweight when his BMI is between 25 and 30 kg/m^2 and obese when his BMI is higher than 30 kg/m^2. We exclude extreme BMI's values from the sample following the Tukey boxplot procedure ($\pm 1.5 \times \text{interquartile range}$). Hence, individuals who have a BMI lower than 12.8 kg/m^2 and higher than 40.6 kg/m^2 are dropped.

2.3. Methodological issues

The main methodological issue consists in treating potential sample selection bias due to the focus on employees and the endogenous relationship between bodyweight and hourly earnings. To correct both problems simultaneously, we bootstrap a three-step parametric model that combines a selection regression and a two-stage least square (2SLS) regression. Formulated by Mroz (1987), this model has been frequently used by labour economists (e.g. Renders, Gaeremynck, and Sercu 2010), but also by health and development economists in order to estimate the effect of body-mass on wages (Shimokawa 2008). Since the objective of

[7] We also produce additional estimates that include more information on the activity (type of contract, Mexican classification of occupations) in the earning function. The results are similar (not shown).

the study is to determine the functional form of the relationship between body-mass and hourly earnings, both linear and quadratic specifications are employed.

2.3.1. Dealing with sample selection bias

About 60% of individuals have a paid activity in the working-age sample (15-65 years old). Thus, the non-random selection of employees might indeed lead to a sample selection bias insofar as various factors explain why an individual works or not, and why an individual works as an employee (44,5%) or a self-employed worker (12,5%) when he works (Nordman and Roubaud 2009). A procedure implemented by Bourguignon, Fournier, and Gurgand (2007) provides a way of correcting for non-randomly selected samples when the first-stage selection equation takes several modalities (inactive, unemployed or unpaid worker; employee; self-employed; entrepreneur). The selection process is based on a *multinomial logit regression* that estimates the probability of belonging to an occupational category in relation to the three other alternatives.

$$P(occupation_{it} = j|X_{it}, Q_{it}) \quad (2)$$

Where *occupation* is the occupational category *j* of an individual *i* at the period *t*. *j* takes four possible alternatives, *j={1,2,3,4}*: inactive, unemployed and unpaid workers; employees; self-employed workers; employers. The reference category is the first alternative *j=1*. Q_{it} represents a set of additional exogenous variables that theoretically affect labour market selection, but are not directly correlated with hourly earnings. As recommended by Nordman and Roubaud (2009), we include in Q_{it} the respondent's household status compared to the self-proclaimed household head (head of household, spouse, child, parent, other) and the household's dependency ratio (number of inactive, unemployed and unpaid individuals, divided by the number of individuals in the household). Moreover, we add to Q_{it} the square of the cognitive skills' score, assuming that the relationship between mental abilities and employment probability is quadratic.[8] Nevertheless, note that the relationship between mental abilities and labour income is assumed linear and not quadratic. For this reason, we only integrate the linear form of the cognitive skills' score into the set of control variables (X_{it}) which structure the earning function.

From this *multinomial logit regression*, four correctional terms (called $\widehat{S_{it}}$) are estimated using the selection method of Bourguignon, Fournier, and Gurgand (2007).[9] Then, these four correctional terms are introduced in the 2SLS regression where the strategy for controlling endogeneity problems is employed.

2.3.2. Correcting for endogeneity problems

Another methodological challenge comes from the endogenous relationship between body-mass and earnings (Cawley, 2004). First, the error component might contain unobserved factors that explain anthropometric and economic status simultaneously, for instance, certain genetic and environmental factors of individuals (physical and socio-cognitive abilities, preferences, etc.). Second, this relationship is potentially affected by the presence of reverse causality. Indeed, weight might influence earnings but it is well known that income, and more generally socioeconomic status, determines anthropometric health (Levasseur 2015). Different strategies for correcting these endogeneity problems have been implemented in the health economics literature. At the beginning, researchers used to run ordinary least square (OLS) regressions explaining current earnings by lagged measures of bodyweight (Sargent and Blanchflower 1994). Although this approach allows the reverse correlation problem to be dealt with, the other factor of endogeneity might persist. Unobserved factors (genes and environment) may still explain lagged weight and current earnings simultaneously. More

[8] It is well known that mental abilities increase employment probability, but we assume that excessive cognitive skills are associated with inactivity or unemployment. Indeed, the "inactive, unemployed and unpaid workers" group is very heterogeneous. Part of this group consists of individuals with high mental skills, such as students, highly educated unemployed, annuitants and also sabbatical workers.

[9] Estimates from the *multinomial logit regression* are available from the authors on request.

recently, academics have therefore implemented a differentiating strategy based on fixed-effects models using longitudinal surveys (Cawley, 2004) or the existence of another individual with highly correlated genes, such as the anthropometric status of parents, children, same-sex siblings or twins who live separately (Baum and Ford 2004; Johar and Katayama 2012). Nevertheless, although fixed-effects strategy eliminates time-invariant heterogeneity influencing both weight and income (particularly genes), this procedure does not deal with unobserved factors that vary over time. Moreover, fixed-effects estimators provide inconsistent results when panel data are based on few waves, as it is the case in our data. Therefore, in line with Cawley (2004), we implement an instrumental variables (IV) strategy based on a random-effects two-stage least squares (2SLS) estimator. In the 2SLS model both the IV equation (3) and structural earning equation (4) take the following form (Angrist and Pischke 2008):

$$B_{it} = \alpha X_{it} + \pi \widehat{S_{it}} + \gamma Z_{it} + \varepsilon_{it} \quad (3)$$

$$\begin{cases} \log Y_{it} = \beta X_{it} + \varphi \widehat{S_{it}} + \rho \widehat{B_{it}} + \mu_{it} \\ \quad with \ \mu_{it} = \left[\vartheta_{it} + \rho (B_{it} - \widehat{B_{it}})\right] \end{cases} \quad (4)$$

Where $\widehat{S_{it}}$ constitutes the vector of correctional terms estimated by the method of Bourguignon, Fournier, and Gurgand (2007) in the selection regression. The inclusion of these terms in the 2SLS regression allows the sample selection bias to be dealt with. Concerning the instrument Z_{it}, it must be a non-weak predictor of the endogenous variable B_{it} (body-mass) conditional on X_{it} (control variables) and must satisfy the exclusion restriction assumption (i.e. not to be related directly to the error component in the structural wage equation (4), $Cov[Z_{it}, \mu_{it}] = 0$). Cawley (2004) uses the BMI of a sibling as instrument, but the data on a sibling relationship across households are not available in the MxFLS.

Economists often use regional data as instruments for endogenous explanatory variables appearing in individual-level equations. Obviously, other regional variables that affect the dependent variable should be controlled for (Wooldridge, 2010, p.95). For instance, in the case of developing countries, Schultz (2003), Kedir (2008) and Shimokawa (2008) instrument anthropometric status with food market prices in the community in order to explain individual earnings. Nevertheless, food market prices in the community appear to be a weak instrument in the Mexican context (results not shown). Accordingly, as Morris (2007) and Pan et al. (2013), we instrument the individual BMI by the median BMI in the municipality in which the respondent lives. Municipalities are the second-level administrative division in Mexico (the first-level is the State). Each municipality has a council that is responsible for providing all amenities for its population (there are 2,438 municipalities in the country).[10] The median BMI is calculated from the active and inactive adult individuals (15-65 years old), excluding the individual *i*. When the individual BMI is analysed in a quadratic way, the square of the median BMI in the municipality is incorporated into the set of instrumental variables.[11]

Theoretically, two principal behavioural links may explain why the area body-mass constitutes a strong instrument of individual BMI (Morris 2007; Pan et al. 2013). First, environmental characteristics (availability of supermarkets, fast-foods, sports field, etc.) could affect food intake and physical activity of community members. It is commonly accepted in the literature that an obesogenic environment can positively influence the body-mass (Swinburn et al. 2011). Even if little is known about the effects of area obesity in Mexico, Villa-Caballero et al. (2006) and Ortiz-Hernández and Janssen (2014) show that neighbourhood characteristics (socioeconomic status, social disorder, etc.) can affect individual body-mass. Second, social factors could strengthen the positive relationship between individual and area BMI. As previously discussed in the conceptual framework

[10] We assume that the number of respondents is large enough to be representative of each municipality (at least 100 adult respondents by municipality).
[11] We test the sensitivity of the results using another instrument that identifies area body-mass: the prevalence of obesity in the municipality. As the results related to the use of this alternative instrument are quite similar, we do not show them in this article. These results are available from the authors on request.

(scenario 3), it is not unreasonable to assume that being overweight might be considered as the physical norm where the majority of the local population is overweight or obese. One must recall the findings in psychology that identify an increasing social acceptance of overweight in the U.S. communities where obesity rates are the highest (Robinson and Christiansen, 2014).

Table 1: The impact of area body-mass on individual BMI: IV equation (3)

	Employees			**Self-employed workers**		
	Linear specification	*Quadratic specification*		*Linear specification*	*Quadratic specification*	
	Individual BMI	Individual BMI	Individual BMI square	Individual BMI	Individual BMI	Individual BMI square
Municipal median BMI	0.745***	1.622*	51.04	0.744***	4.563**	243.9**
	(19.75)	(1.770)	(1.052)	(8.181)	(2.307)	(2.411)
Municipal median BMI square		-0.0165	-0.184		-0.0730*	-3.863**
		(-0.948)	(-0.199)		(-1.925)	(-1.989)
Done replications/100	100	100	100	100	100	100
Unique individuals	11170	11170	11170	2149	2149	2149
R-squared	0.184	0.184	0.169	0.129	0.131	0.123
F-statistic	134.91	130.12	119.98	19.97	19.41	19.04
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: (1) These estimates come from the first-stage of 2SLS estimates.
(2) All the control variables (X_{it}) and the correctional terms (S_{it}^{j}) are included.
(3) The standard errors are corrected bootstrapping the three-step parametric model (z-statistics are in parentheses).
Significance levels of coefficients: *** p<0.01; ** p<0.05; * p<0.1.
Source: MxFLS (2002-2012).

Empirical estimates presented in Table 1 suggest that the area body-mass satisfies the first requirement of an instrument. As expected, even after controlling for all covariates, the median BMI in the municipality is a significant and positive predictor of individual BMI. Besides, it is comforting to note that all first-stage F-statistics on the excluded instrument are relatively high. In addition, Angrist and Pischke (2008) recommend estimating a reduced-form model, by regressing the dependent variable on the instrumental variable (Table 2). The instrument is said strong if its effect on the dependent variable (earnings) is proportional to the effect of the endogenous variable (individual BMI) on the same dependent variable. Therefore, it will be important to compare the magnitude of the coefficients and their significance between the reduced-form model and the IV model. The results in Table 2 are encouraging. As clearly expressed by Angrist and Pischke (2008, p.213), if one can perceive the causal relation of interest in the reduced-form regression, it probably indicates that causality does exist. In other words, in Table 2, the magnitude of the instrument coefficient and his significance would be due to the existence of a causal relationship between individual BMI and hourly earnings, assuming that the exclusion restriction hypothesis is respected.

Table 2: The impact of area body-mass on log-hourly earnings: reduced-form model

	Employees		**Self-employed workers**	
	Linear specification	*Quadratic specification*	*Linear specification*	*Quadratic specification*
Municipal median BMI	0.0241***	0.488***	0.0577***	1.310***
	(4.349)	(3.221)	(4.334)	(3.877)
Municipal median BMI square		-0.00875***		-0.0240***
		(-3.072)		(-3.689)
Done replications/100	100	100	100	100
Unique individuals	11170	11170	2149	2149
R-squared	0.224	0.224	0.135	0.140
F-statistic	185.52	177.71	21.20	21.70
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: (1) All the control variables (X_{it}) and the correctional terms (S_{it}^{j}) are included.
(2) The standard errors are corrected bootstrapping the three-step parametric model (z-statistics are in parentheses). Significance levels of coefficients: *** p<0.01; ** p<0.05; * p<0.1.
Source: MxFLS (2002-2012).

In classical IV models, the second requirement of an instrument (exclusion restriction assumption) cannot be tested empirically (Wooldridge, 2010). In our application, this means that the median BMI in the municipality should not directly correlate with individual earnings through channels other than the individual BMI. It means that the link between municipal median BMI and individual income must go through individual BMI. Such assumption could be violated if there are unobserved factors that determine both area body-mass and individual earnings (e.g. the area development level). To neutralise this possible effect, Morris (2007) recommends controlling for the level of development of the municipality. Hence, we included in the set of control variables the infrastructural development index of the municipality. Although it is not possible to prove the exclusion restriction assumption, it may be informative to analyse the correlation between residuals μ_{it} from the structural earning equation (4) and the instrument Z_{it} (Shimokawa, 2008). If the instrument is correlated with the error term from the earning equation (4), it might cast doubt on the validity of the selected instrument. Nonetheless, Figure A.1 of the Appendix does not report any significant correlation between residuals and the area median BMI. Therefore, we assume that the area body-mass constitutes a good instrument of individual BMI conditional on the set of selected covariates.

Finally, a potential limit in the use of a municipality-scale instrument comes from the lack of a sampling weight which makes the survey community-representative. Moreover, the instrument is constructed within-sample. Thus, we attempt to take the sample variance related to the instrumentation procedure into account bootstrapping the standard errors of the whole model (Wooldridge, 2010).

2.4. A complementary semiparametric analysis

Shimokawa (2008) proposes to test the robustness of results from the 2SLS model using a semiparametric partially linear estimator. The less-restrictive form of this model allows the effect of individual BMI on log-hourly wages to be visualised clearly in a plot without being limited by the linearity assumption. In this alternative model, we control for endogeneity problems using the control function method formulated by Blundell and Powell (2003). The control function approach consists in introducing estimated residuals from the IV equation (3) into the structural earning equation, as follows:

$$\log Y_{it} = \beta X_{it} + \varphi \widehat{S_{it}} + F(B_{it}) + \rho \widehat{\varepsilon_{it}} + \mu_{it} \qquad (5)$$

Where *F(.)* is an unknown function insofar as there is not restrictive linear assumption on the relationship between BMI (B_{it}) and log-hourly earnings ($log\, Y_{it}$). As in the previous parametric model, we correct standard errors bootstrapping the three steps simultaneously. $\widehat{\varepsilon_{it}}$ is the estimated error term from the IV equation (3), namely the control function. Significant coefficients of $\widehat{\varepsilon_{it}}$ (i.e. $\hat{\rho}$) could indicate the presence of endogeneity and so might justify the control function method used. This model is said to be semi-parametric because the control variables (X_{it}), the correctional terms of sample selection ($\widehat{S_{it}}$) and the control function ($\widehat{\varepsilon_{it}}$) are regressed parametrically, while the coefficient of B_{it} is estimated from a non-parametric Gaussian kernel function. More formally, the partially linear model is based on the following Robinson's (1988) double residual estimator:

$$logY_{it} - E(logY_{it}|B_{it}) = \beta[X_{it}, \widehat{S_{it}} - E(X_{it}, \widehat{S_{it}}|B_{it})] + \rho[\widehat{\varepsilon_{it}} - E(\widehat{\varepsilon_{it}}|B_{it})] + \mu_{it} \quad (6)$$

The conditional means are non-parametrically estimated using univariate Gaussian Kernel regressions and coefficients β and ρ are estimated parametrically using pooled OLS. Finally, the function $F(B_{it})$ is obtained by graphing:

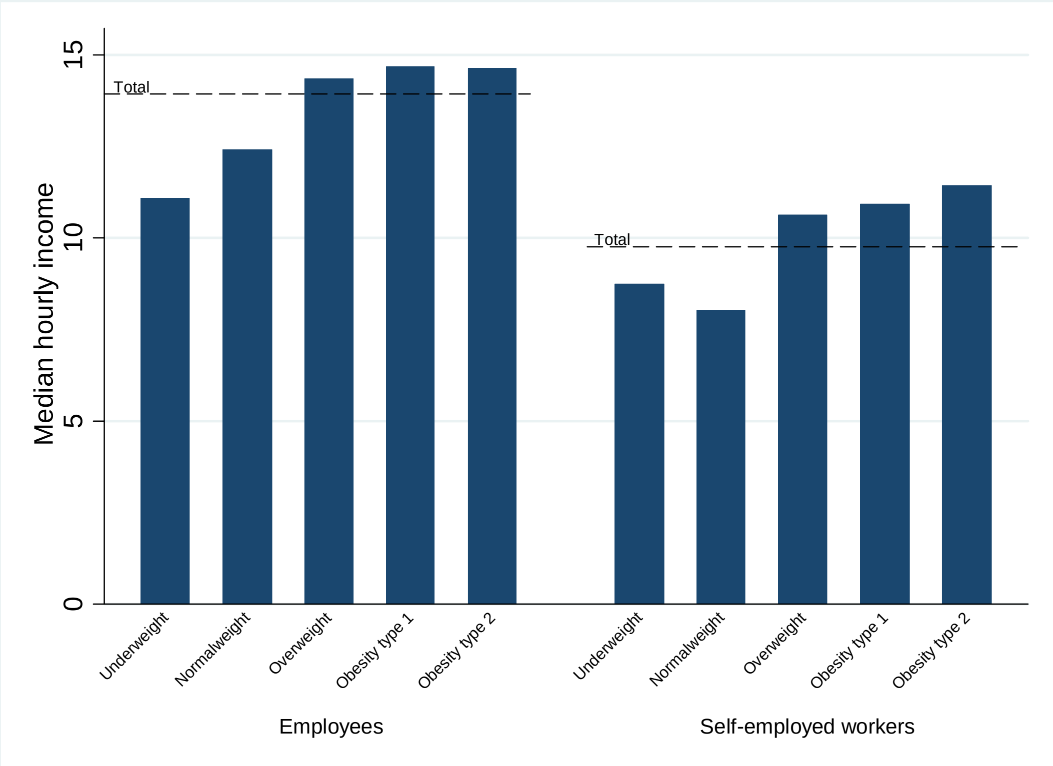
$$\begin{cases} E(logY_{it}|B_{it}) - \hat{\beta}E(X_{it}, \widehat{S_{it}}|B_{it}) - \hat{\rho}E(\widehat{\varepsilon_{it}}|B_{it}) \\ \text{against } B_{it} \end{cases} \quad (7)$$

3. RESULTS

Figure 2 shows the hourly income gap between employees and self-employed workers. The median salary is about 13.7 pesos per hour for employees, while half of self-employed workers accumulate at the most 9.6 pesos for one hour worked. Descriptive statistics presented in Table A.1 of the Appendix outline that employees have better living conditions than self-employed workers, revealing better economic independence, education, health and access to public services (measured by the infrastructural development index). Moreover, it is interesting to note in Figure 1 that hourly earnings tend to increase with the body-mass category, both for employees and self-employed workers, at least up to overweight status. Indeed, the hourly income gap between overweight and obese workers is not significant.

Despite the apparent presence of a link between individual body-mass and hourly earnings, it would be premature to conclude on a direct and positive causal relationship between both factors. We speculate that other dimensions can influence and alter the relationship; hence the importance of implementing an IV strategy.

Figure 2: Median hourly income by body-mass category



Source: MxFLS (2002-2012).

3.1. Parametric analysis

To explore the effects of BMI on log-hourly income in Mexico, we run a bootstrapped three-step parametric model with 100 replications (i.e. multinomial selection regression followed by a 2SLS regression), using both linear and quadratic specifications. As discussed earlier, an IV strategy is particularly justified and area body-mass constitutes the better option to instrument individual BMI given the context of the study.

The results from the parametric analysis are presented in Tables 3 and 4 for employees and self-employed workers, respectively. For both samples, three models are executed: (i) an OLS regression without correction of selection bias and endogeneity (equation 1); (ii) a two-stage model that only corrects for the sample selection bias (see Bourguignon, Fournier and Gurgand, 2007); (iii) a three-step IV model in which the individual BMI is instrumented by the median BMI of the municipality (equation 4). In each model, the relationship between

BMI and log-hourly earnings is specified in its linear and quadratic form. If we refer to the R^2, the explanatory power of the model is relatively high, especially for the employees' sample (explains 22% of the total variance).

The influence of control variables that structure the earning function allows our results to be compared with the literature in labour economics and thus to justify the relevance of the model. As shown in Tables 3 and 4, the workforce is better paid in areas with relatively high occupational opportunities (North and Central regions, large cities), particularly in the case of self-employment. Besides, as expected, the effects of education and cognitive skills on earnings are twice higher for employees than for self-employed workers (Gong, Soest and Villagomez, 2004). Moreover, the quadratic influence of age on hourly earnings reflects the accumulation of experience and the ageing process that accompany the worker's life cycle. Note that income inequalities against women are slightly higher in self-employment than in salaried employment. Finally, employees in a couple (married or common-law relationship) are significantly better paid than their single, separated or widowed counterparts (for additional explanations, see Chun and Lee, 2001).

In OLS estimates, when BMI is specified using a linear form, the relationship between BMI and hourly earnings follows a positive trend (Tables 3 and 4, columns 1). Note that this positive trend is stronger for self-employed workers (Table 2) than for employees (Table 1). The results from the three-step IV model confirm this positive trend (column 5), although the magnitude of coefficients varies between OLS and three-step IV estimates[12]. Nevertheless, the municipal median BMI coefficients from the reduced-form model (Table A.3 of the Appendix) are proportional to the individual BMI coefficients from the structural earning equation (Table 1 and 2). Consequently, we can conclude that we really do observe a causal

[12] The coefficient of the individual BMI is almost seven times larger in IV estimates (column 5) than in OLS estimates (column 1). Reasons for this gap will be discussed in the part 3.2).

relationship between individual BMI and hourly income (Angrist and Pischke 2008).

Table 3: Results from the parametric analysis, employees' sample

Dependent: log-hourly income	**OLS**		**OLS with selection equation**		**Three-step IV model**	
	Linear	*Quadratic*	*Linear*	*Quadratic*	*Linear*	*Quadratic*
Column Number	*1*	*2*	*3*	*4*	*5*	*6*
BMI	0.00548***	0.0192	0.00499***	0.0177	0.0376***	0.831**
	(3.597)	(1.492)	(3.282)	(1.378)	(4.010)	(2.495)
BMI square		-0.000249		-0.000230		-0.0143**
		(-1.071)		(-0.993)		(-2.393)
Age	0.0339***	0.0335***	0.0290***	0.0286***	0.0131*	-0.0142
	(9.316)	(9.108)	(6.437)	(6.318)	(1.907)	(-0.967)
Age square	-0.00032***	-0.00031***	-0.00020***	-0.00020***	-4.99e-05	0.000236
	(-6.569)	(-6.417)	(-3.671)	(-3.581)	(-0.604)	(1.475)
Gender (male)	0.0769***	0.0759***	0.109***	0.108***	0.110***	0.0645*
	(5.159)	(5.086)	(5.528)	(5.488)	(4.803)	(1.857)
Cognitive skills' score	0.00192***	0.00192***	0.00114***	0.00114***	0.00123***	0.00116**
	(6.274)	(6.268)	(3.430)	(3.426)	(3.202)	(2.126)
Years of schooling	0.0564***	0.0564***	0.0445***	0.0445***	0.0459***	0.0456***
	(32.89)	(32.87)	(17.91)	(17.91)	(14.25)	(13.31)
Marital status (couple)	0.0883***	0.0878***	0.119***	0.119***	0.0915***	0.0508*
	(5.535)	(5.507)	(6.324)	(6.293)	(4.021)	(1.748)
Children number	-0.0129***	-0.0130***	-0.0108**	-0.0109**	-0.00765	-0.00983
	(-2.781)	(-2.789)	(-2.330)	(-2.340)	(-1.289)	(-1.482)
Infrastructural development index	0.0233***	0.0233***	0.00884	0.00883	0.00487	0.00340
	(3.745)	(3.739)	(1.331)	(1.329)	(0.608)	(0.363)
Region (South-East)	-0.185***	-0.185***	-0.0585**	-0.0591**	-0.0678**	-0.110***
	(-10.22)	(-10.26)	(-2.241)	(-2.267)	(-2.225)	(-2.847)
Large city	0.168***	0.168***	0.111***	0.111***	0.119***	0.124***
	(9.740)	(9.729)	(5.550)	(5.553)	(4.909)	(4.826)
Middle-sized city	0.0741***	0.0739***	-0.00204	-0.00206	0.000475	-0.000482
	(2.960)	(2.954)	(-0.0747)	(-0.0755)	(0.0146)	(-0.0142)
Small city	0.103***	0.102***	0.0600***	0.0596***	0.0627**	0.0395
	(4.607)	(4.586)	(2.595)	(2.579)	(2.556)	(1.220)
2006' survey	0.0684***	0.0686***	0.0103	0.0106	0.0213	0.0420
	(4.161)	(4.174)	(0.532)	(0.548)	(0.938)	(1.642)
2012' survey	0.0615***	0.0618***	0.0175	0.0178	0.00669	0.0239
	(4.039)	(4.056)	(1.058)	(1.076)	(0.342)	(1.005)
Correctional term 1			-0.0437	-0.0456	-0.0675	-0.183
(inactive & others)			(-0.512)	(-0.533)	(-0.855)	(-1.515)
Correctional term 2			-0.134**	-0.134**	-0.128**	-0.124*
(employee)			(-2.419)	(-2.417)	(-2.393)	(-1.838)
Correctional term 3			1.452***	1.448***	1.185***	0.860**
(self-employment)			(5.665)	(5.648)	(3.879)	(2.544)
Correctional term 4			-1.819***	-1.821***	-1.724***	-1.828***
(entrepreneur)			(-4.721)	(-4.726)	(-3.342)	(-3.130)
Constant	0.772***	0.598***	1.277***	1.115***	0.691***	-9.483**
	(11.20)	(3.436)	(11.32)	(5.657)	(3.016)	(-2.222)
Done replications/100					100	100
Unique individuals	11170	11170	11170	11170	11170	11170
R^2	0.218	0.218	0.222	0.222		

Notes: (1) The municipal median BMI is used as instrumental variable and median BMI square is added as instrumental variable in quadratic specifications.
(2) The standard errors are corrected bootstrapping the three-step IV model (z-statistics are in parentheses).
Significance levels of coefficients: *** p<0.01; ** p<0.05; * p<0.1.

Source: MxFLS (2002-2012).

Table 4: Results from the parametric analysis, self-employed workers' sample

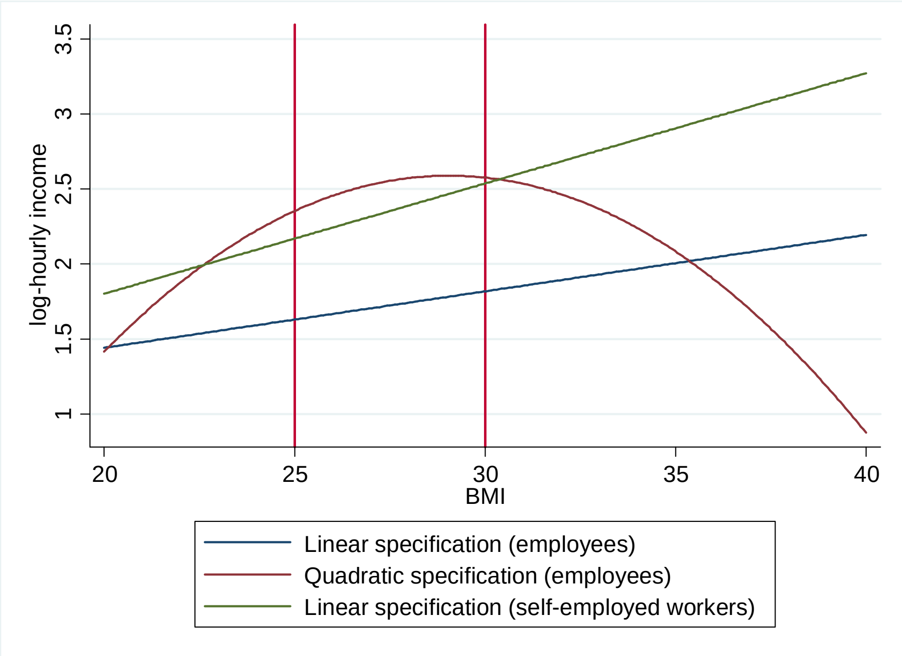
Dependent: log-hourly income	**OLS**		**OLS with selection equation**		**Three-step IV model**	
	Linear	*Quadratic*	*Linear*	*Quadratic*	*Linear*	*Linear*
Column Number	*1*	*2*	*3*	*4*	*5*	*6*
BMI	0.00989**	0.00511	0.01000**	0.00615	0.0735***	3.600
	(2.417)	(0.145)	(2.431)	(0.174)	(2.844)	(0.646)
BMI square		8.53e-05		6.82e-05		-0.0631
		(0.137)		(0.109)		(-0.630)
Age	0.0216**	0.0217**	0.0238*	0.0239*	-0.00571	-0.108
	(2.123)	(2.132)	(1.675)	(1.682)	(-0.281)	(-0.623)
Age square	-0.000241**	-0.000243**	-0.000255*	-0.000257*	2.78e-05	0.00115
	(-1.981)	(-1.989)	(-1.726)	(-1.731)	(0.141)	(0.627)
Gender (male)	0.137***	0.138***	0.142**	0.142**	0.158**	-0.0672
	(3.038)	(3.037)	(2.292)	(2.295)	(2.316)	(-0.138)
Cognitive skills' score	0.00133*	0.00133*	0.00129	0.00130	0.00120	-0.00393
	(1.665)	(1.661)	(1.357)	(1.358)	(1.193)	(-0.477)
Years of schooling	0.0240***	0.0241***	0.0230***	0.0230***	0.0250***	0.000730
	(5.198)	(5.204)	(2.854)	(2.853)	(2.885)	(0.0187)
Marital status (couple)	0.0950**	0.0949**	0.101*	0.101*	0.0412	0.0779
	(2.006)	(2.003)	(1.731)	(1.732)	(0.514)	(0.278)
Children number	-0.0269**	-0.0270**	-0.0265*	-0.0264*	-0.0187	-0.0360
	(-1.982)	(-1.988)	(-1.926)	(-1.926)	(-1.378)	(-0.475)
Infrastructural development index	0.0150	0.0151	0.0141	0.0142	0.00477	-0.0688
	(0.880)	(0.883)	(0.698)	(0.702)	(0.203)	(-0.540)
Region (South-East)	-0.227***	-0.228***	-0.222***	-0.222***	-0.254***	-0.105
	(-5.445)	(-5.446)	(-2.685)	(-2.687)	(-2.679)	(-0.295)
Large city	0.346***	0.346***	0.337***	0.337***	0.325***	0.333
	(6.675)	(6.676)	(5.435)	(5.435)	(4.799)	(1.260)
Middle-sized city	0.206***	0.206***	0.199**	0.199**	0.176*	0.0251
	(2.704)	(2.712)	(2.289)	(2.293)	(1.922)	(0.0539)
Small city	0.145***	0.146***	0.143**	0.143**	0.100*	-0.0620
	(2.684)	(2.691)	(2.503)	(2.504)	(1.678)	(-0.204)
2006' survey	0.0665	0.0658	0.0590	0.0590	0.0891	0.0720
	(1.482)	(1.465)	(1.028)	(1.028)	(1.290)	(0.237)
2012' survey	0.248***	0.248***	0.245***	0.245***	0.232***	0.275
	(5.790)	(5.780)	(4.943)	(4.943)	(4.355)	(0.981)
Correctional term 1			0.155	0.154	0.345	0.106
(inactive & others)			(0.532)	(0.530)	(1.050)	(0.0785)
Correctional term 2			0.113	0.112	0.378	-0.132
(employee)			(0.189)	(0.188)	(0.538)	(-0.0518)
Correctional term 3			0.0915	0.0907	-0.0462	0.333
(self-employment)			(0.416)	(0.411)	(-0.168)	(0.340)
Correctional term 4			0.0957	0.103	0.360	-3.044
(entrepreneur)			(0.0754)	(0.0810)	(0.267)	(-0.577)
Constant	1.013***	1.073**	0.958	1.007	0.333	-45.86
	(4.826)	(2.170)	(1.204)	(1.092)	(0.342)	(-0.633)
Done replications/100					100	87
Unique individuals	2149	2149	2149	2149	2149	2149
R^2	0.127	0.127	0.127	0.127		

Notes: (1) The municipal median BMI is used as instrumental variable and median BMI square is added as instrumental variable in quadratic specifications.
(2) The standard errors are corrected bootstrapping the three-step IV model (z-statistics are in parentheses).
Significance levels of coefficients: *** p<0.01; ** p<0.05; * p<0.1.

Source: MxFLS (2002-2012).

Causal relationship between individual BMI and earnings appears to be different between employees and self-employed workers. For employees, although the linear effect of BMI on wages is slightly positive, the quadratic specification shows relevant findings and highlights the presence of a turning point from which this effect becomes negative. Figure 2 illustrates the main results from the parametric analysis. The inflection point, from which the effect of BMI on employees' wages passes from positive to negative, is around 29 kg/m² (near to obesity status). By contrast, in the case of self-employed workers, the linear specification is much more relevant than the quadratic specification, no significant turning point being found (Table 4). Basing on the linear specification, Figure 2 reveals a linear and positive effect of BMI on the self-employed workers' earnings.

Figure 3: Effect of BMI on hourly income from the parametric analysis



Note: The red lines refer to the thresholds between normal-weight, overweight, and obesity.
Source: MxFLS (2002-2012).

Obviously, findings from parametric analysis are simplistic, since the functional form of the relation is predetermined (linear or quadratic). However, the causal relationship between body-mass and labour income might be more complex. Hence, we complete the study with a semi-parametric analysis in which the functional form of the link between BMI and earnings is not explicitly specified.

3.2. Robustness test using a semiparametric model

Estimates from the structural wage equation (5) are displayed in Table A.2 of the Appendix.

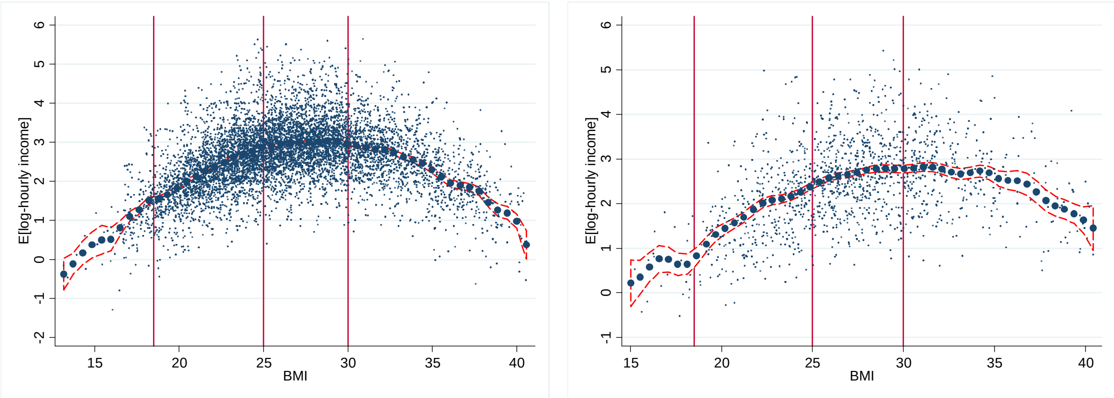
As in the previous parametric model, the three-step semiparametric model (sample selection, IV and structural earning equations) is bootstrapped, but we run only 50 replications because the semiparametric partially linear estimator is highly computationally intensive.

In Table A.2 of the Appendix, the significance of the control function's coefficient, $\hat{\rho}$ (the coefficient of estimated residuals from equation 3), identifies the presence of endogeneity in the relationship between individual BMI and hourly earnings (Wooldridge, 2010). Such unobserved heterogeneity can explain why, in Tables 3 and 4, OLS coefficients (columns 1 and 2) are smaller than instrumented coefficients (columns 5 and 6). When $\hat{\rho}$ is negative (positive), it means that unobserved factors underestimate the positive (negative) effect of individual BMI on hourly earnings (before implementing the IV strategy). In other words, referring to Table A.2 of the Appendix, two types of unobserved factors lead to biased estimates: (i) factors underestimating the positive effect of a relatively moderate BMI (normal-weight and overweight in the Mexican case) on hourly earnings (e.g. socio-cognitive skills, risky consumption); (ii) factors underestimating the negative impact of a high BMI (obesity) on hourly earnings (e.g. factors related to professional ambition such as stress and lack of free time). Hence, we introduce a quadratic control function in the structural earning equation (5) in order to correct factors that tend to underestimate the negative effects of a high BMI on hourly income. In practice, we instrument individual BMI and its square by the municipal median BMI and its square.

For employees, Figure A.2(a) in the Appendix only depicts a slight positive effect. By contrast, introduction of a quadratic control function appears much more relevant since this procedure controls the unobservable heterogeneity tending to underestimate the negative effect of a relatively high BMI (obesity) on hourly earnings. As shown in Figure 3(a), and in line with the parametric analysis, the causal relationship between BMI and hourly wage takes a U-inverted form, with an inflection point at 29 kg/m^2.

Figure 4: Effect of BMI on hourly income from the semiparametric analysis (using a quadratic control function)

a) Employees *b) Self-employed workers*



Note: The red lines refer to the thresholds between normal weight, overweight, and obesity.
Source: MxFLS (2002-2012).

In the case of self-employed workers, a linear control function seems to be more appropriate than a quadratic control function (according to the significance of $\hat{\rho}$ in Table A.2 of the Appendix). This result is consistent with the parametric analysis: we do not find any significant turning point between individual BMI and hourly income for self-employed workers (Figure A.2(b)). In other words, the effect of BMI on hourly earnings seems to be positive and linear. However, we observe a decline in this positive effect beyond a BMI of 32 kg/m² in Figure A.2(b) of the Appendix. In addition, when we use a quadratic control function for self-employed workers, the causal relationship between BMI and hourly earnings, presented in Figure 3(b), is strong and positive between 18.5 kg/m² and 26 kg/m², then relatively flat between 26 kg/m² and 32 kg/m², and becomes negative beyond a BMI of 32 kg/m². Nonetheless, few self-employed workers have a BMI greater than 35 kg/m² in our sample (only 121 observations overpass this level). Thus, the variance artificially increases (confidence interval in red dashes). This lack of observations at the right-tail of the distribution may explain why the quadratic specification in the parametric analysis does not identify any significant inflection point for self-employed workers (Table 4). To sum up, it seems clear that overweight status leads to a higher income in self-employment. However, we fail to assess the presence of a significant income gap between overweight self-employed workers and their obese counterparts, given the small number of obese individuals with a BMI higher than 35 kg/m² in this sample.

3.3. The effect of BMI on labour force participation

Probit estimates in Table 5 (column 2) shows a U-inverted shape relationship between individual BMI and labour force participation. However, the Wald test leads us to reject the assumption of exogeneity between both factors. It is the reason why we implement additional *IV-Probit* estimations in which the individual BMI is instrumented by the median BMI of the municipality. Controlling for endogeneity, Table 5 (column 4) suggests that the quadratic trend in the relationship between BMI and employment probability disappears: the individual BMI affects labour force participation in a linear and negative fashion. Column 3 of Table 5 indicates that, on average, one extra kg/m² decreases the employment probability by almost 6 percentage points. This result does not appear so surprising given the low rate of underweight in Mexico (Table A.1 of the Appendix).

Table 5: Marginal effects of individual BMI on labour force participation

	***Probit* estimator**		***IV-Probit* estimator**	
	Linear specification	*Quadratic specification*	*Linear specification*	*Quadratic specification*
BMI	-0.00317*	0.0872***	-0.0596***	-0.27
	(-1.684)	(5.206)	(-5.544)	(-0.73)
BMI square		-0.00161***		0.00386
		(-5.429)		(0.572)
Observations	34392	34392	34392	34392
Pseudo R²	0.3431	0.3438		
Wald test of exogeneity (Chi²)			29.19	31.16
(p-value)			(0.000)	(0.000)

Notes: (1) All the control variables (X_{it}) are included.
(2) The municipal median BMI is used as instrumental variable and median BMI square is added as instrumental variable in quadratic specifications.
(3) Panel data are pooled. Significance levels of coefficients (z-statistics are in parentheses): *** p<0.01; ** p<0.05; * p<0.1.

Source: MxFLS (2002-2012).

4. DISCUSSION

Our findings emphasise the complex labour market consequences of the nutritional situation that characterises emerging economies with high obesity prevalence. Using panel data from the *Mexican Family Life Survey* (MxFLS), the main purpose of this study was to estimate the effects of body-mass index (BMI) on hourly earnings in Mexico, focusing on the employees and self-employed workers. First, we implemented a bootstrapped three-step parametric model, based on an expanded Mincer earning function, to control for potential sample selection bias and endogeneity problems. We instrumented the individual BMI using the median BMI of the municipality. Both linear and quadratic specifications were analysed. Second, we ran a bootstrapped three-step semi-parametric model in order to precisely determine the nature of the relationship and test the robustness of previous parametric findings.

Assuming the instrumental variables strategy implemented in the study is valid, we discover the presence of a causal relationship between individual body-mass and individual hourly income in Mexico. As observed by Shimokawa (2008) in China, the relationship between BMI and hourly wage of employees is characterised by a U-inverted form. However, the turning point of this curve differs from the Chinese case. While in China, the reversal point is around the normal-weight category (BMI close to 22-23 kg/m²), this point revolves around a BMI of 29 kg/m² in Mexico (close to obesity status). In other words, in Mexico, overweight employees have better earnings than thinner and fatter employees. Such divergences between China and Mexico are probably due to sociocultural aspects resulting from differences in the BMI distribution. While in China, the obesity rate does not exceed 10%, one-third of Mexicans are obese. In fact, overweight status has recently become the physical norm in Mexico, and probably a social norm too. Such social acceptance of overweight would explain why overweight is somewhat preferred, or at least accepted, in salaried employment.

To summarize, obesity is significantly penalized in salaried jobs in Mexico, while overweight is commonly accepted and probably rewarded in some occupational categories. This hybrid situation highlights two paradoxical social phenomena that are occurring simultaneously: (i) the increasing social rejection of obesity, due to the large diffusion and appropriation of thinness ideals (scenario 2); (ii) the social acceptance of overweight, due to past nutritional deprivations (scenario 1) and the recent generalised weight gain (scenario 3). Consequently, as expected, the causal relationship between body-mass and wages in Mexico takes the form of a right-leaning U-inverted curve, due to the possible concomitance of scenarios 1, 2 and 3 as described in the conceptual framework.

Similar to coalminers in India (Dinda et al., 2006) and self-employed workers in Guinea (Glick and Sahn, 1998), overweight does not lead to income penalties for self-employed workers in Mexico. The parametric and semi-parametric estimates indicate a linear and positive effect of BMI on hourly earnings, at least up to a certain threshold (when the BMI reaches 32 kg/m²). However, lack of observations at the right-tail of the distribution (BMI higher than 35 kg/m²) significantly alters our estimates and does not allow us to conclude whether there is (or whether there is no) a turning point in case of self-employed workers. Nevertheless, even if this turning point does exist, it would only occur from a higher BMI than 32 kg/m², a higher BMI level than the level which characterizes the employees (29 kg/m²).

The absence of income penalties for self-employed obese (at least for obese of level 1) is not particularly surprising in the case of Mexico. First, self-employment is not exposed to discriminatory behaviours from employers and co-workers. Moreover, since self-employment is largely composed of vulnerable and uneducated workers who work informally (Gong, Soest and Villagomez, 2004), overweight may be considered (by clients and/or the individual himself) as an asset allowing to better perform physical and manual labours (scenario 1).

These different findings between employees and self-employed workers emphasise the presence of sociocultural heterogeneity within the Mexican society in the way weight is perceived and stigmatised. This is similar to the case of Hispanic and African-American women in the United States for whom obesity does not lead to any wage penalties (Cawley, 2004). Additional analyses are required to explore potential sociocultural heterogeneity which influences the causal relationship between body-mass and earnings within a country, comparing different sub-samples (specific to gender, living area, occupational category or observation time). Furthermore, other case studies based on emerging countries with different nutritional contexts (e.g. India, South-Africa, Egypt, Brazil) should be carried out in order to identify sociocultural heterogeneity worldwide.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflict of interest.

APPENDIX

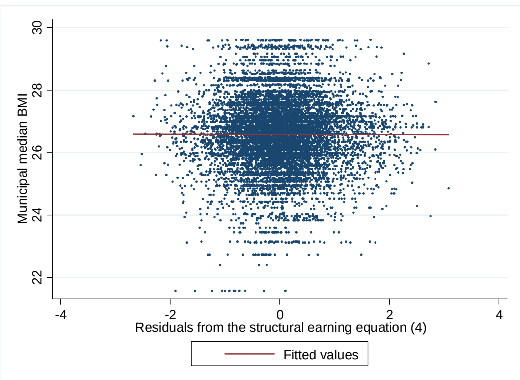
Table A.1: Descriptive statistics

	Whole sample		Employees' sample		Self-employed workers' sample	
	Mean	S-D	Mean	S-D	Mean	S-D
Inactive, unemployed and unpaid workers	0.40					
Employees	0.44					
Self-employed workers	0.13					
Employer	0.03					
log-hourly income			2.67	0.76	2.38	0.87
Household head	0.29		0.42		0.65	
Spouse of the head	0.25		0.13		0.17	
Child of the head	0.39		0.40		0.15	
Parent of the head	0.05		0.05		0.03	
Others (cousin, brother-in-law, etc.)	0.01		0.00		0.01	
Economic dependency ratio	0.43	0.20	0.35	0.19	0.39	0.20
Age	33.32	13.98	32.82	12.16	41.55	12.16
Gender (male)	0.46		0.67		0.72	
Cognitive skills' score	50.44	24.91	52.46	24.02	45.35	24.71
Years of schooling	10.97	4.79	11.76	4.66	9.44	4.67
Marital status (couple)	0.58		0.58		0.79	
Number of children	1.49	1.45	1.49	1.42	1.55	1.44
Infrastructural development index	3.14	1.24	3.26	1.24	3.10	1.25
Region (South-East)	0.18		0.17		0.24	
Large city (>100000 inhbts)	0.38		0.44		0.34	
Middle-sized city (15000-100000 inhbts)	0.09		0.09		0.09	
Small city (2500-15000 inhbts)	0.11		0.11		0.13	
Rural area (<2500 inhbts)	0.42		0.36		0.45	
2002's survey	0.40		0.35		0.50	
2006's survey	0.28		0.30		0.22	
2012's survey	0.32		0.35		0.29	
BMI	26.43	4.91	26.41	4.60	27.45	4.44
Underweight (BMI<18.5)	0.03		0.02		0.01	
Normal-weight (18.5>=BMI>25)	0.39		0.39		0.28	
Overweight (25>=BMI>30)	0.35		0.39		0.43	
Obesity (BMI>35)	0.23		0.20		0.27	

Source: MxFLS (2002-2012).

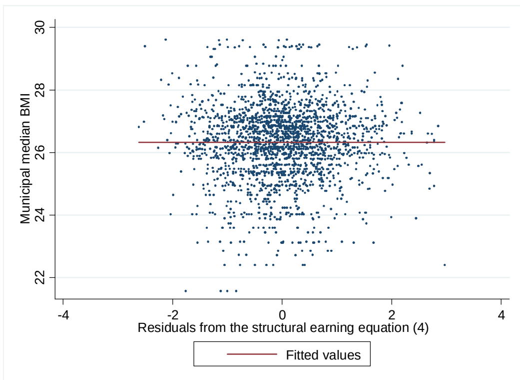
Figure A.1: Correlation between municipal median BMI and residuals from the structural earning equation (4)

a) Employees



Note: Fitted values are nonsignificant.
Source: MxFLS (2002-2012).

b) Self-employed workers



Note: Fitted values are nonsignificant.
Source: MxFLS (2002-2012).

Table A.2: Structural earning equation (5) from the semiparametric analysis

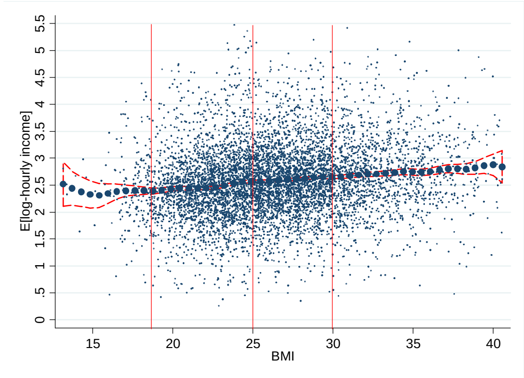
	Employees		Self-employed workers	
	Linear instrumentation	Quadratic instrumentation	Linear instrumentation	Quadratic instrumentation
Estimated residuals ($\hat{\rho}$) 1	-0.0340***	-0.756***	-0.0627***	-2.324
	(-2.978)	(-3.048)	(-2.697)	(-1.534)
Estimated residuals ($\hat{\rho}$) 2		0.0130***		0.0401
		(2.913)		(1.481)
Age	0.0111	-0.0126	-0.00772	-0.0984*
	(1.534)	(-1.135)	(-0.393)	(-1.696)
Age square	-2.45e-05	0.000219*	4.63e-05	0.00100
	(-0.318)	(1.935)	(0.222)	(1.642)
Gender (male)	0.108***	0.0669**	0.165**	0.0493
	(4.876)	(2.001)	(2.554)	(0.374)
Cognitive skills' score	0.00121***	0.00114**	0.000992	-0.00364
	(3.743)	(2.364)	(1.209)	(-1.037)
Years of schooling	0.0449***	0.0455***	0.0234***	0.00135
	(15.79)	(12.29)	(2.928)	(0.0824)
Marital status (couple)	0.0934***	0.0533*	0.0337	0.0131
	(4.992)	(1.797)	(0.510)	(0.121)
Children number	-0.00694	-0.00959	-0.0197	-0.0348
	(-1.296)	(-1.613)	(-1.268)	(-1.473)
Infrastructural development	0.00407	0.00331	0.00261	-0.0627
	(0.483)	(0.412)	(0.115)	(-1.198)
Region (south)	-0.0619*	-0.106***	-0.236***	-0.0762
	(-1.907)	(-3.140)	(-2.736)	(-0.520)
Large city size	0.112***	0.121***	0.318***	0.307***
	(4.491)	(6.111)	(4.437)	(3.075)
Middle-sized city	-0.00432	-0.000434	0.155	-0.0391
	(-0.135)	(-0.0109)	(1.606)	(-0.201)
Small city size	0.0604**	0.0406	0.0858	-0.104
	(2.006)	(1.264)	(1.315)	(-0.702)
2005-06 survey	0.0185	0.0397	0.0874	0.0872
	(0.873)	(1.478)	(1.547)	(0.898)
2010-12 survey	0.00475	0.0227	0.230***	0.237***
	(0.230)	(0.950)	(4.977)	(2.918)
Correctional term 1	-0.0652	-0.172	0.377	0.639
(inactive)	(-0.651)	(-1.269)	(1.180)	(1.318)
Correctional term 2	-0.142**	-0.125*	0.441	0.692
(employee)	(-2.453)	(-1.785)	(0.698)	(0.730)
Correctional term 3	1.295***	0.901**	0.0115	0.488
(self-employment)	(4.425)	(2.502)	(0.0590)	(1.051)
Correctional term 4	-1.843***	-1.849***	-0.0451	-4.096
(entrepreneur)	(-4.000)	(-3.281)	(-0.0375)	(-1.393)
Done replications/50	50	50	50	50
Unique individuals	11170	11170	2149	2149
R-squared	0.210	0.210	0.122	0.125

Notes: (1) The area median BMI is used as instrumental variable in the linear control function procedure. When the control function is quadratic, the equation of instrumental variables (4) splits into two regressions. The term "estimated residuals 1" is extracted from an OLS regression where individual BMI is instrumented by municipal median BMI and its square. Then, the term "estimated residuals 2" is extracted from a second OLS regression where individual BMI square is instrumented by municipal median BMI and its square.
(2) The standard errors are corrected bootstrapping the three-step semiparametric model (z-statistics are in parentheses). Significance levels of coefficients: *** p<0.01; ** p<0.05; * p<0.1.

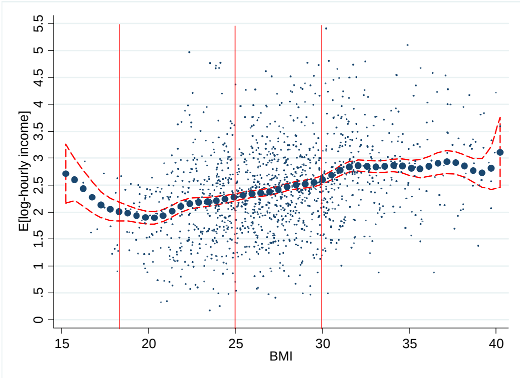
Source: MxFLS (2002-2012).

Figure A.2: Effect of BMI on hourly income from the semiparametric analysis (using a linear control function)

a) Employees



b) Self-employed workers



Note: The red lines refer to the thresholds between normal weight, overweight, and obesity.
Source: MxFLS (2002-2012).

REFERENCES

ANGRIST Joshua D., **PISCHKE** Jörn-Steffen, 2008, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 393 p.

BAUM Charles L., **FORD** William F., 2004, "The wage effects of obesity: a longitudinal study", *Health Economics*, 13(9), pp. 885–899.

BLUNDELL Richard, **POWELL** James L., 2003, "Endogeneity in Nonparametric and Semiparametric Regression Models", *in Advances in Economics and Econometrics*, Cambridge University Press, Econometric Society Monographs.

BONNEFOND Céline, **CLÉMENT** Matthieu, 2014, "Social class and body weight among Chinese urban adults: The role of the middle classes in the nutrition transition", *Social science & medicine (1982)*, 112C, pp. 22–29.

BOURGUIGNON François, **FOURNIER** Martin, **GURGAND** Marc, 2007, "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons", *Journal of Economic Surveys*, 21(1), pp. 174–205.

BREWIS Alexandra, 2003, "Biocultural aspects of obesity in young Mexican schoolchildren", *American Journal of Human Biology*, 15(3), pp. 446–460.

BREWIS Alexandra A., **WUTICH** Amber, **FALLETTA-COWDEN** Ashlan, **RODRIGUEZ-SOTO** Isa, 2011, "Body Norms and Fat Stigma in Global Perspective", *Current Anthropology*, 52(2), pp. 269–276.

BRUNELLO Giorgio, **D'HOMBRES** Béatrice, 2007, "Does body weight affect wages?: Evidence from Europe", *Economics & Human Biology*, 5(1), pp. 1–19.

CAMPOS-VAZQUEZ Raymundo M., **LÓPEZ-CALVA** Luis Felipe, **LUSTIG** Nora, 2016, "Declining Wages for College-Educated Workers in Mexico: Are Younger or Older Cohorts Hurt the Most?", SSRN Scholarly Paper, ID 2723691, Rochester, NY, Social Science Research Network.

CAWLEY John, 2004, "The Impact of Obesity on Wages", *Journal of Human Resources*, 39(2), pp. 451–474.

CHUN H, **LEE** I, 2001, "Why do married men earn more: productivity or marriage selection?", *Economic Inquiry*, 39(2), pp. 307–319.

CLASSEN Tim, 2017, "Changes Over Time in the Relationship of Obesity to Education Accumulation", *Eastern Economic Journal*, 43(3), pp. 496–519.

COLCHERO M. Arantxa, **BISHAI** David, 2012, "Weight and earnings among childbearing women in Metropolitan Cebu, Philippines (1983-2002)", *Economics and Human Biology*, 10(3), pp. 256–263.

DINDA Soumyananda, **GANGOPADHYAY** P. K., **CHATTOPADHYAY** B. P., **SAIYED** H. N., **PAL** M., **BHARATI** P., 2006, "Height, weight and earnings among coalminers in India", *Economics & Human Biology*, 4(3), pp. 342–350.

GLICK Peter, **SAHN** David E., 1998, "Health and productivity in a heterogeneous urban labour market", *Applied Economics*, 30(2), pp. 203–216.

GONG Xiaodong, **SOEST** Arthur van, **VILLAGOMEZ** Elizabeth, 2004, "Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico", *Economic Development and Cultural Change*, 53(1), pp. 1–36.

JOHAR Meliyanni, **KATAYAMA** Hajime, 2012, "Quantile regression analysis of body mass and wages", *Health Economics*, 21(5), pp. 597–611.

KEDIR Abbi M., 2008, "Health and Wages: Panel Evidence on Men and Women using IV Quantile Regression", Discussion Papers in Economics, 08/37, Department of Economics, University of Leicester.

LEVASSEUR Pierre, 2015, "Causal effects of socioeconomic status on central adiposity risks: Evidence using panel data from urban Mexico", *Social Science & Medicine*, 136–137, pp. 165–174.

MINCER Jacob A., 1974, "Schooling, Experience, and Earnings", NBER Books, National Bureau of Economic Research, Inc.

MORRIS Stephen, 2007, "The impact of obesity on employment", *Labour Economics*, 14(3), pp. 413–433.

MROZ Thomas A., 1987, "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions", *Econometrica*, 55(4), pp. 765–799.

MULATU Mesfin Samuel, **SCHOOLER** Carmi, 2002, "Causal connections between socio-economic status and health: reciprocal effects and mediating mechanisms", *Journal of Health and Social Behavior*, 43(1), pp. 22–41.

NORDMAN Christophe J., **ROUBAUD** François, 2009, "Reassessing the Gender Wage Gap in Madagascar: Does Labor Force Attachment Really Matter?", *Economic Development and Cultural Change*, 57(4), pp. 785–808.

ORTIZ-HERNÁNDEZ Luis, **JANSSEN** Ian, 2014, "Social disorder, physical activity and adiposity in Mexican adults: Evidence from a longitudinal study", *Health & Place*, 30, pp. 13–19.

PAN Jay, **QIN** Xuezheng, **LIU** Gordon G., 2013, "The impact of body size on urban employment: Evidence from China", *China Economic Review*, 27, pp. 249–263.

POULAIN Jean-Pierre, 2017, *The Sociology of Food: Eating and the Place of Food in Society*, Bloomsbury Publishing, 312 p.

RAVEN John, 2000, "The Raven's Progressive Matrices: Change and Stability over Culture and Time", *Cognitive Psychology*, 41(1), pp. 1–48.

RENDERS Annelies, **GAEREMYNCK** Ann, **SERCU** Piet, 2010, "Corporate-Governance Ratings and Company Performance: A Cross-European Study", *Corporate Governance: An International Review*, 18(2), pp. 87–106.

RENZAHO André M. N, 2004, "Fat, rich and beautiful: changing socio-cultural paradigms associated with obesity risk, nutritional status and refugee children from sub-Saharan Africa", *Health & Place*, 10(1), pp. 105–113.

ROBINSON Eric, **CHRISTIANSEN** Paul, 2014, "The changing face of obesity: exposure to and acceptance of obesity", *Obesity (Silver Spring, Md.)*, 22(5), pp. 1380–1386.

ROBINSON Peter M., 1988, "Root- N-Consistent Semiparametric Regression", *Econometrica*, 56(4), pp. 931–954.

RUNGE C. Ford, 2007, "Economic consequences of the obese", *Diabetes*, 56(11), pp. 2668–2672.

SARGENT J. D., **BLANCHFLOWER** D. G., 1994, "Obesity and stature in adolescence and earnings in young adulthood. Analysis of a British birth cohort", *Archives of Pediatrics & Adolescent Medicine*, 148(7), pp. 681–687.

SARLIO-LAHTEENKORVA Sirpa, **SILVENTOINEN** Karri, **LAHELMA** Eero, 2004, "Relative Weight and Income at Different Levels of Socioeconomic Status", *American Journal of Public Health*, 94(3), pp. 468–472.

SCHULTZ T. P., 2003, "Wage rentals for reproducible human capital: evidence from Ghana and the Ivory Coast", *Economics and Human Biology*, 1(3), pp. 331–366.

SHIMOKAWA Satoru, 2008, "The labour market impact of body weight in China: a semiparametric analysis", *Applied Economics*, 40(8), pp. 949–968.

SWINBURN Boyd A., **SACKS** Gary, **HALL** Kevin D., **MCPHERSON** Klim, **FINEGOOD** Diane T., **MOODIE** Marjory L., **GORTMAKER** Steven L., 2011, "The global obesity pandemic: shaped by global drivers and local environments", *Lancet (London, England)*, 378(9793), pp. 804–814.

VILLA-CABALLERO Leonel, **CABALLERO-SOLANO** Víctor, **CHAVARRÍA-GAMBOA** Marcelo, **LINARES-LOMELI** Patricia, **TORRES-VALENCIA** Erendira, **MEDINA-SANTILLÁN** Roberto, **PALINKAS** Lawrence A, 2006, "Obesity and socioeconomic status in children of Tijuana", *American journal of preventive medicine*, 30(3), pp. 197–203.

VOGL Tom S., 2014, "Height, skills, and labor market outcomes in Mexico", *Journal of Development Economics*, 107, pp. 84–96.

WHO, 2000, "Obesity: preventing and managing the global epidemic", *WHO Technical Report*, 894, pp. 5–37.

WOOLDRIDGE Jeffrey M., 2010, *Econometric Analysis of Cross Section and Panel Data*, Second Edition, Cambridge MA, The MIT Press.