(Non-Elderly) Health Insurance Coverage, Moral Hazard and Crime

Nicola Ciccarelli*
Tilburg University

January 21, 2015

Abstract

This paper studies the effect of changes in health insurance coverage on criminal activity in the U.S. using a county level dataset for the period 2005-2010. Using various Panel Data Instrumental Variables estimators we find that an increase in the health insurance coverage implies an increase in the number of per-capita larcenies and car thefts while violent crimes are not affected by changes in insurance coverage. We connect our findings to the vast literature on health insurance gain and moral hazard [e.g., Finkelstein et al. (2012)] and to the literature which relates changes in (disposable) incomes and criminal activity.

Keywords: health insurance coverage, property crime, violent crime.
JEL Classification: G22, I11, I18, K14, K42.

"The amount of crime is determined not only by the rationality and preferences of would-be criminals but also by the economic and social environment created by public policies." [Becker (1993)]

1 Introduction

The purpose of the present study is to investigate the relationship between the non-elderly health insurance coverage and crime using county-level data for the U.S. between 2005 and 2010.[1] Becoming eligible for public health insurance (i.e., Medicaid/CHIP) or acquiring private health insurance for an individual that was uninsured implies a large increase in health care utilization and in health care expenditure [Finkelstein et al. (2012), Meer and Rosen (2004), Hadley et al. (2008), Jerant et al. (2013), Kreider et al. (2009), Ward et al. (2007), Dong (2000)] since the health insurance gain is characterized by ex-post moral hazard (i.e., increased use of health care because of shared health care costs) [Jerant et al. (2013)].[2] The increase in health care expenditure is associated with a decrease of disposable income for the (average) individual who is gaining health insurance, which in turn might be related to a higher probability of committing property crimes (i.e., larceny theft, car theft, burglary, arson) for the same individual.[3] As a matter of fact, a negative relationship between real incomes (or real wages) and the probability of committing property crimes has been found in the previous literature [Fagan and Freeman (1999), pp. 225-228; Raphael and Winter-Ebmer (2001); Gould et al. (2002); Machin et al. (2004); Mocan and Bali (2010)][4]. In our context, the insurance gain for the "average

*CONTACT DETAILS: Room K631, PO Box 90153, 5000 LE Tilburg. Phone: +31 13 466 3254. Email: n.ciccarelli@tilburguniversity.edu

[1] For simplicity we do not specify that we use the non-elderly health insurance coverage, but we simply refer to health insurance coverage throughout the text.

[2] Arrow (1968): "Mr. Pauly has a very interesting sentence: "The above analysis shows, however, that the response of seeking more medical care with insurance than in its absence is a result not of moral perfidy, but of rational economic behavior.""

[3] In the case that the public health insurance (i.e. Medicaid/CHIP programs) coverage increases in a state, the state and the federal government must increase taxes or decrease funding for other services such as educational services [Emanuel et al. (2004)], which in turn will have negative effects on the disposable income of the (average) individual. Since the health insurers need to recover their costs (i.e., the benefits paid to the insurees), to recover the administrative costs, and to make a profit, with the term "average individual" we mean the individual that contributes to the profit of the insurer, that is, the individual that pays a premium which is larger than the expected benefit (received from the insurer). The average individual is the individual that has average health status, thus an individual that is neither in perfect health status nor in a deadly condition (i.e., an individual affected by a tumor or AIDS).

[4] For instance, Fagan and Freeman (1999) report that "the growth of illegal income opportunities in the 1980s, in a context of declining wages and job losses in inner cities, increased the potential returns from crime and perhaps altered the basic economic calculus for many young people".

individual" implies a higher (total) profit for the insurer and a lower disposable income for the insuree, that is, a lower real income compared to the case that the individual did not buy health insurance. Additionally, the composition of the health insurance coverage by income levels might play a role. Approximately 94% of high income earners are covered by health insurance, 81.8% of middle income earners are covered by health insurance while only 56.7% of low income earners are covered by health insurance in 2007 [Holahan (2010), Exhibit 4], thus it is highly likely that an increase in the insurance coverage in a county is caused by an increase in insurance coverage of low or middle income earners. In turn, the net loss caused by the insurance gain (i.e., the net loss is the premium minus the expected benefit for the insuree) has stronger negative effects on incomes of poor/middle class individuals than on incomes of high earners.

We also study the relationship between the insurance coverage and violent crimes (i.e., robbery, murder, rape, aggravated assault). Robbery is classified as a violent crime even though the goal of a robbery is not violence *per se*, but is a pecuniary gain. We hypothesize that an increase in the insurance coverage might increase or might have no effect on the number of robberies since robbery is much riskier than larceny theft, thus a criminal might prefer larceny theft to robbery. For the remaining violent crimes (i.e., murder, rape, aggravated assault) we hypothesize that increases in the insurance coverage might have no effect[5] or a negative effect on violent crime rates. An increase of the health insurance coverage in a community (i.e., county) implies better health of individuals in the community [Kasper et al. (2000)], which in turn is correlated with higher social capital of the community [Carpiano (2007), pp. 650-651][6] Finally, Sampson et al. (1997) [Table 4-5] find that a higher level of social capital implies a lower number of (one type of) violent crimes.[7]

Several issues arise when analyzing the causal effect of health insurance coverage on crime rates. The coverage rate in a county might be affected by the crime level of the county (*reverse causality bias*). After a crime shock in a county, firms might relocate to safer counties, and since most of the private insurance is co-paid by the employee and the employer (i.e., employer-based insurance), the insurance coverage might change.[8] Further biases may be caused by *omitted variables* that are correlated with both the insurance coverage and crime, and by measurement error of insurance coverage data.[9] We cope with these problems using various techniques. Firstly, we use the Within Group estimator to control for county fixed effects which may be correlated with both the insurance coverage and crime rates. In addition, year fixed effects are included to control for nationwide confounding shocks, and we use a set of socio-economic controls by county and the number of police forces in $t-1$ by county in order to decrease the probability of omitted variable bias.[10] Secondly, we check whether our estimates suffer from omitted variable bias using the Altonji et al. (2005)'s method for omitted variable bias. Thirdly, we tackle all the endogeneity biases using various types of IV estimators (i.e the FE-IV estimator, FD-IV estimator, the (static) Difference GMM estimator) and using different sets of instruments. Following Murray (2006), the use of different sets of instruments -i.e., instruments characterized by different rationales- tend to improve the robustness of the results, in the case that the various (second-stage) IV estimates point in the same direction. Finally, if the insurance coverage is *measured with error*, but the instruments are valid (i.e., exogenous with respect to the error term and correlated with the instrumented variable), then the estimate of the effect of the insurance coverage on crime will be consistent[11].

The remainder of the paper is organized as follows. Section 2 presents the literature review and the

[5]The drop of the disposable income (after the insurance gain) might not incentivize to do more violence. In addition, the main determinants of violent crimes are non-pecuniary: 1) drug shocks [Levitt (2004)], 2) exposure to violent movies [Dahl et al. (2009)], 3) jail peer effects [Bayer et al. (2009)], 4) abortion [Donohue and Levitt (2001)], 5) changes in police forces and in incarceration rates [Levitt (2004)], etc.

[6]Carpiano (2007) [pp. 650-651]: "Poor health may restrict certain residents from engaging others in the neighborhood and accessing differential forms of social capital (e.g. Klinenberg (2001)).

[7]The term "collective efficacy" is used in Sampson et al. (1997) as a synonym of social capital.

[8]Also rich families tend to relocate to safer counties after a crime shock [Cullen and Levitt (1996), page 4], which in turn might have an effect on the insurance coverage.

[9]Several published studies such as Ericson et al. (2012) and Miller (2012) use the SAHIE (Small Area Health Insurance Estimates) dataset, thus the measurement error bias might be less of a concern.

[10]Finally, we allow the error to be heteroskedastic and correlated within panels with the use of clustered standard errors at the county level.

[11]The estimate will be consistent even in the case that the instrumental variable is measured with error but the measurement error of the instrument(s) is (are) uncorrelated with the measurement error of the instrumented variable.

microeconomic model of crime. Section 3 describes the data. Section 4 describes the methods used in this study. Section 5 presents the main empirical results of the study. Section 6 shows the presence of (ex-post) moral hazard. In Section 7 we check whether the increase of the insurance coverage has other behavioral effects. Section 8 concludes.

2 The Determinants of (Property) Crime

We present a review of the crime literature focusing on the determinants of property crime in Section 2A. We present a microeconomic model of crime in Section 2B; using this model we obtain that a decrease of disposable income implies an increase of property crime rates, which evidence is in line with Section 2A.

2A Literature Review

In this section we review the main determinants of property crime.

(Disposable) Income and Crime

Several papers study the effect of income on crime. Imrohoroglu, Merlo and Rupert (2004) construct a (structural) dynamic equilibrium model of crime which is dependent on disposable income (i.e., incomes minus a government tax on income) [Imrohoroglu, Merlo and Rupert (2004), p. 711] and on other variables and they estimate the model using U.S. data for 1980-1996. The disposable income is one of the main determinants of crime [Imrohoroglu, Merlo and Rupert (2004), Table 7]; simulating the model, the authors find that the increase in average (disposable) incomes that was observed in 1980-1996 would decrease the property crime rate by 20% in 1996, *ceteris paribus*.[12]

Other papers find a negative relationship between incomes and crime. Differently from Imrohoroglu, Merlo and Rupert (2004), the following papers are focused on incomes, not on disposable incomes. Mocan and Bali (2010) use the 1958 Philadelphia Birth Cohort, an individual panel of 27,160 persons from 1970 to 1984, and the authors find that increases in the real per-capita income implies a decrease in larceny theft, both in the case that the dependent variable is a binary variable indicating whether an individual committed a crime [Mocan and Bali (2010), Table 4][13] and when the dependent variable is the number of crimes committed by an individual [Mocan and Bali (2010), Table 5]. Using state-level data for the U.S. in 1971-1997, Raphael and Winter-Ebmer (2001) find that increases in per-worker incomes imply a strong decrease of property crimes [Raphael and Winter-Ebmer (2001), Table 2].[14]

Wages and crime

Gould et al. (2002) study the effect of wages and unemployment on crime rates in the U.S. using IV techniques. The authors find that an increase in wages (or a decrease in the unemployment rate) implies a decrease of property and violent crime rates. Also, if the wages of the non-college educated individuals increase, then the decrease of crime rates is even stronger. Men with low education have the highest probability of committing crimes [Gould et al. (2002), pp. 57], which in turn explains the previous finding. Machin et al. (2004) study the relationship between wages and property crime in the UK during 1975-1996. Firstly, areas characterized by low wages for the bottom 25th percentile of the wage distribution show a high amount of crime. Secondly, using statistical techniques, the authors find that increases in the wages of the 25th percentile of the wage

[12] Imrohoroglu, Merlo and Rupert (2004) does not analyze the effect of changes in the income's tax on crime; anyhow, an increase in the income's tax imply a decrease of the disposable income, which in turn implies a higher level of crime. An individual that was uninsured in period t and that buys health insurance in period $t+1$ will pay a premium (in $t+1$) such that the premium is equal to the health care expenditure in year t plus the profit (and the administrative cost) of the insurer in $t+1$ plus an increase in health care expenditure in $t+1$ (i.e., the moral hazard effect); if we think about the latter two parts (i.e., the profit and the moral hazard effect) as a tax on income (or a tax on health care expenditure), then (using Imrohoroglu, Merlo and Rupert (2004)) an increase in insurance coverage increases the property crime rate.

[13] In this case, an increase in the real per-capita income also implies an increase of burglaries and robberies [Mocan and Bali (2010), Table 4].

[14] Per-worker incomes are found dividing the personal incomes by the percentage of employed individuals in the state. Anyhow, also using per-capita incomes, Raphael and Winter-Ebmer (2001) find the same effect on the number of property crimes.

distribution imply a decrease in the amount of property crimes. Therefore, wages and even more the wages of the poorest individuals in the U.K. are associated with property crime rates. Western et al. (2006) [Table 5] find that increases in earnings have a strong negative effect on the probability of being arrested for both Black and White males. Additionally, the authors find that a 100$ increase in median weekly earnings is estimated to halve the risk of prison admission in the U.S.. Narayan and Smith (2004) use Granger causality tests to understand the relationship between various types of crime and real male average weekly earnings and unemployment in the case of Australia. The authors find that robbery or serious assault are not cointegrated with real male average weekly earnings.

2B A Microeconomic Model of Criminal Behavior

This section is based on the model of Gronau (1977) where an individual allocates time between work, home production and leisure. We modify the model of Gronau (1977) and we adapt it to the case where an individual allocates time between work, crime and leisure. Several papers show that the majority of criminal offenders had legitimate activities before incarceration [Ehrlick (1973), Fagan and Freeman (1999), Grogger (1998) and others],[15] thus a model of criminal activity must take into account that an individual optimizes with respect to crime, work and leisure.

Let there be a single-person household. The person maximizes the utility U, which is dependent on a combination of goods (X) and leisure (L).

$$U = U(X, L). \tag{1}$$

The goods can either be purchased in the market (X_M) or they can be stolen from other individuals (X_C), but the composition of X does not affect U. I measure the value of the crime goods (X_C) in terms of their market equivalents (i.e., the cost of the good in the market). Let X_M denote market expenditures; then, the total consumption is composed of market goods consumption and of crime goods consumption,

$$X = X_M + X_C. \tag{2}$$

Crime goods are a function of time in crime (C) and of government expenditure in deterrence (e.g., police forces, CCTV cameras), D,[16]

$$X_C = f(C, D), \tag{3}$$

subject to decreasing marginal productivity ($f' > 0, f" < 0$). We simplify our notation for the returns to crime function as $X_C = f(C)$[17]. The decreasing marginal productivity of criminal activity is caused by rational choice: at first, criminals perpetrate the crimes with the highest returns, then criminals might take into consideration crimes with lower returns.

The maximization of U is bound by the budget constraint and the time constraint. The *endogenous budget constraint* takes the form in (4),

$$X_M = WN + V - Benefits(p) - p \times NetCost \tag{4}$$

where W is the (constant) wage rate, N is the number of (labor) market working hours, V is the exogenous initial income, $Benefits(p)$ is the health care expenditure which is a function of the probability of having

[15] For example, Ehrlick (1973) [page 528, footnote 10] reports that "studies of prisoners in federal, state, and local correctional institutions in the United States show that a majority of these offenders did have legitimate occupational experience -mainly in unskilled occupations- prior to their apprehension, and that only a small fraction never worked (see Ehrlick (1975)).

[16] We explicitly take into account the probability of arrest (a) in the returns to crime function. The probability of arrest (a) is a function of time spent in crime (C) and of the (exogenous) government deterrence expenditure (D), thus $a = g(C, D)$. Since crime goods are function of time in crime and of the probability of arrest (i.e., $X_C = f(C, a)$), using $a = g(C, D)$, we obtain that $X_C = f(C, g(C, D))$, which can be simplified to $X_C = f(C, D)$.

[17] [...] since our main goal is not the study of the effect of deterrence on crime.

health insurance p (with $p = 0$ or $p = 1$)[18], $NetCost = [Premium - Benefits]$ is the net cost of insurance for a single individual.[19] [20] The time constraint is[21]

$$L + C + N = T \tag{5}$$

The maximization of equation (1) subject to the constraints (4) and (5) is equivalent to maximizing the Lagrangian function $\mathcal{L} = U((X_M + f(C)), L) + \lambda(WN + V - Benefits(p) - pNetCost - X_M) + \mu(T - L - C - N)$ with respect to L, C, N, X_M. The necessary conditions for an *interior solution*[22] call for the marginal product of crime (f') to equal the marginal rate of substitution between goods and leisure $\left(\frac{\partial U/\partial L}{\partial U/\partial X}\right)$, which equals the shadow price of time ($\frac{\mu}{\lambda}$), and if the individual works in the market ($N > 0$), f' is also equal to the wage rate:[23]

$$\frac{\partial U/\partial L}{\partial U/\partial X} = f' = \frac{\mu}{\lambda} = W. \tag{6}$$

If the person does not work in the market ($N = 0$), then the last equality in equation (6) might not hold [see Appendix B], and in this case the necessary conditions are[24]

$$\frac{\partial U/\partial L}{\partial U/\partial X} = f' = \frac{\mu}{\lambda}. \tag{7}$$

Analysis of the Model

There are several combinations of working hours, crime and leisure based on the preferences of the individual, the wage and the marginal productivity of crime (f'); for instance, an individual can split the total time (T) into time in criminal activity (C) and in leisure time (L) without working in the labor market (i.e., $N = 0$). We analyze the possible combinations in Appendix C, and we present the comparative statics in the following discussion.

Comparative Statics

In Figure 2-1, the individual has discretion over T hours of time and the exogenous income is equal to the distance TD.[25] The returns to crime ($f(C)$) are decreasing and are given by the curve $TDB'C$. The more time the individual spends in crime (as measured by the horizontal distance from point T), the greater is the amount of crime goods that are produced. If the individual spends all her time in crime, she can produce an amount of OC units of goods. In the absence of work opportunities, the curve $TDB'C$ is the opportunity frontier

[18][...] since the health insurance gain implies higher health care expenditure (i.e., moral hazard), that is, $[Benefits(p)|p = 1] > [Benefits(p)|p = 0]$. See Section 6 where we show the presence of moral hazard.

[19]The total net cost of (private) health insurance is calculated as the difference between incurred premiums earned and benefits paid for (private) health insurance. The total net cost of (private) health insurance at the national U.S. level are around 150 billion$ in each year for the period 2005-2010 [Catlin et al. (2007), Catlin et al. (2008), Hartman et al. (2009), Hartman et al. (2010), Martin et al. (2011), Martin et al. (2012)]. Approximately 183 millions of U.S. individuals below 65 (i.e., 70% of U.S. individuals) have (private) health insurance [using Exhibit 4 of Holahan (2010)], thus the individual net cost is around 820$ per year.

[20]Also in the case that the individual was uninsured in $t-1$ and gains public health insurance in t, he will consume more health care but the majority of the cost will be paid by the local and federal governments through higher taxes and budget cuts [Emanuel et al. (2004)], which might increase the number of property crimes in the community. Only 9% of non-elderly U.S. individuals are covered by public health insurance in 2007 [Holahan (2010)]; we abstract from the public health insurance (only) for our microeconomic model.

[21]Thus, we can rewrite equation (1) as

$$U = U(X, L) = U^{\dagger}(X, L, T - L) = U^{\dagger}(X, L, C + N)$$

i.e., equation (1) implies that C and N are perfect substitutes.

[22]More details on non-interior solutions (e.g., the individual is only working or only doing crime) are reported in Appendix C.

[23]The maximization of the Lagrangian function used to obtain equation (6) and (7) are reported in Appendix B.

[24]The wage rate (W) may be smaller than the marginal productivity of crime ($W < f'$) because the marginal productivity of work might be smaller than the marginal productivity of crime. For example, an individual with very low education is likely to be a low productivity worker, and since wages reflect labor productivity, then the individual receives a low wage, which finally might lead to ($W < f'$). A more detailed explanation of the possible combinations of work, crime and leisure decisions is given below.

[25]D is the initial wealth of the individual which corresponds to $V - Benefits(p) - p \times NetCost$.

which includes all the possible combinations of X and L. On the contrary, if there are work opportunities, the individual can sell her (labor) working time and buy market goods, so that she can expand the opportunity frontier to $TDB'E$. Given the wage rate W (i.e., the slope of the line AE), the individual can supply working hours in exchange of goods along the price line AE. Depending on the leisure preference of the individual, there are two possible combinations of X and L: 1) a goods-intensive combination of X and L, 2) a leisure-intensive combination of X and L. In the first case, the individual chooses a goods-intensive combination of X and L at the optimum (i.e., the point B) where she spends OL time units in leisure, spends LN time units in labor market work, and spends NT time units in crime. In the second case, the individual has a high preference for leisure, choosing as his optimum combination the point B'. In this case she does not work in the market, but splits her time between leisure (OL') and crime ($L'T$).[26]

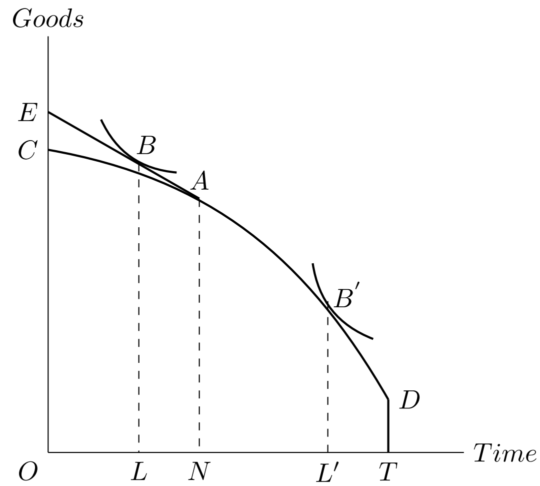


Figure 2-1

To analyze the properties of the model, let us assume that there is a transition from being uninsured ($p = 0$) to being insured ($p = 1$). The effects of the transition are reported in Figure 2-2. The transition implies that the exogenous income $V - Benefits(p) - pNetCost$ decreases for two reasons: 1) after the individual becomes insured, moral hazard implies that the individual increases the health care utilization and expenditure (i.e., $[Benefits(p)|p = 1] > [Benefits(p)|p = 0]$); 2) if the individual buys health insurance, she pays the net cost of insurance which is larger than zero since the health insurer needs to repay the administrative costs and the insurer wants to have a positive profit. A decrease of the exogenous income decreases the amount of market goods that the individual can buy. The change is reflected in a vertical shift of the crime production curve $TDB_0'C_0$ to $TB_1'C_1$. The change does not affect the shape of the curve but only its location, thus the change does not affect the marginal productivity of crime ($f'(C)$) nor it affects the slope of the line E_0A_0 (i.e., the wage). Since W is fixed, there is *no substitution effect* between working and committing crime since the marginal returns of crime ($f'(C)$) and work (W) are left unchanged.[27] As a conclusion, if the individual prefers a goods-intensive combination of X and L (i.e., point B_0 before the transition), and given the pure income effect caused by the decrease in exogenous income, she increases the amount of time in market work from L_0N to L_1N,[28] while the time in crime remains equal to NT.

If the individual prefers a leisure-intensive combination of X and L (i.e., point B_0' before the transition), then the *income effect* caused by the transition incentivize the individual to spend more time in crime (in order

[26] Also, a third possibility exists: if the marginal productivity of crime at zero hours of crime is smaller than the wage ($f'(0) < W$) [see the point D in Figure 2-1], then the individual does not spend time in crime but she enjoys leisure time and she works in the labor market [see Appendix C for more details about this case].

[27] The point N (the point where $f'(C) = W$) is unchanged, thus the time in crime NT does not change.

[28] Thus, her leisure time decreases from OL_0 to OL_1.

to counterbalance the decrease in goods that she can buy). Thus, we end in the point B_1'.[29]

To conclude, if we consider all the individuals in the economy, an increase in the insurance coverage increases the time in crime, leading to a higher crime rate.

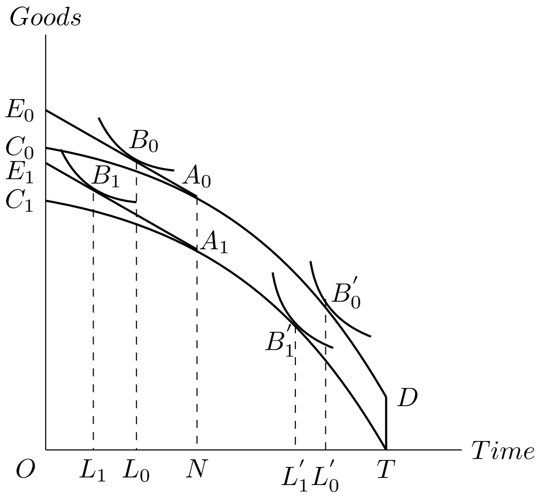


Figure 2-2

3 Data

Our dataset is composed by a panel of 3082 counties in 50 states plus the District of Columbia (i.e., around 98% of U.S. counties) from 2005 to 2010.[30] The dependent variables used in our study are the number of crimes reported per 1000 individuals, and the main explanatory variable is the health insurance coverage, both by county per year. The latter measure is the fraction of individuals that are younger than 65 and that are covered by either public or private health insurance. At the age of 65, all U.S. citizens are covered by public health insurance (i.e. Medicare), thus we explicitly use data for non-elderly individuals. The final dataset is created by merging county-level data from various sources. The panel is almost a balanced panel as shown in Table 1.[31]

[Table 1 about here]

We use county level crime data, including the number of known offenses and arrests, from the NACJD (National Archive of Criminal Justice Data) which gathers and elaborate the FBI's Uniform Crime Reports (UCR) data.[32] County level crime data is created by the NACJD based on police agencies' (real) data or on imputed data.[33] Crimes are divided into property crimes (i.e., larceny theft, car theft, burglary, arson) and violent crimes (i.e., murder, rape, aggravated assault, robbery).[34]

For the county level insurance coverage we use the Small Area Health Insurance Estimates (SAHIE) which contains data on the fraction of individuals between 0 and 64 years old that are covered by either private or public health insurance.[35] The SAHIE dataset contains data for the year 2005 and successive years. Prior to

[29] The time in crime increases from $L_0'T$ to $L_1'T$.

[30] Once the variable for the police forces is added to the model, the number of counties drop to around 2797 counties per year (i.e., a total of 16784 observations) since the FBI data for the police forces contains many missing observations. The other random variables contain few or no missing data.

[31] In this section, we present the descriptive statistics for the data used in Column (4) and Column (8) of Table 4 and of Table 5 [Within Group estimates].

[32] The NACJD data was accessed in June/July 2014. The data is available at `http://www.icpsr.umich.edu/icpsrweb/NACJD/archive.jsp`. We use the "County-Level Detailed Arrest and Offense Data" for the year 2005 (i.e., dataset ICPSR 4717), year 2006 (i.e., dataset ICPSR 23780), year 2007 (i.e., dataset ICPSR 25114), year 2008 (i.e., dataset ICPSR 27644), year 2009 (i.e., dataset ICPSR 30763), year 2010 (i.e., dataset ICPSR 33523).

[33] The methodology for the imputation of the data is described in `http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/33523/version/2`

[34] The detailed FBI definitions of each type of crime are reported at `https://www2.fbi.gov/ucr/cius2009/about/offense_definitions.html`.

[35] The private health insurance can be either the employer-based health insurance or the private non-group heal insurance while the public health

2005, experimental estimates for the insurance coverage are available for a small minority of counties.

Table 2 presents the descriptive statistics for the county level insurance coverage and crimes (normalized per 1000 individuals). 82% of 0-64 years old individuals are covered by health insurance in average and the distribution of the health insurance coverage is highly skewed to the left [Figure 8-3]. Among the property crimes, larceny theft and burglary are the most common crimes; the average number of larceny theft per 1000 individuals is 13.9. Violent crimes are less frequent than property crimes with the exception of aggravated assault: 1.83 aggravated assaults are committed per 1000 individuals in average, and aggravated assault is the third most common crime after larceny theft and burglary.[36]

[Table 2 about here]

An extensive set of controls is used in the analysis in order to avoid -or at least to weaken - the probability of omitted variable bias. Table 3 presents the descriptive statistics for the controls. The data on the size of the police force is obtained from the FBI;[37] we use the sum of sworn police officers and civilian employees in year $(t-1)$ in order to decrease the probability of reverse causality bias. The (time-varying) socio-economic controls are obtained from the U.S. Census Bureau,[38] and the Bureau of Economic Analysis.[39] We control for variables that may be correlated with crime rates and that may be correlated with the insurance coverage: unemployment rate (in %), population density (i.e., population per square mile), poverty rate,[40] real personal income (in 2005 U.S. dollars), the amount of the per-capita social security recipients (i.e., disabled, retired, widowed individuals), eight dummies for race[41] where the baseline is White with non-Hispanic origin, gender, three variables for the age distribution (the fraction of the population between 15 and 19 years old, between 20 and 24 years old, and older than 64, respectively)[42]. Poverty rates, unemployment rates, real incomes, races, the percentage of males in a county, and the geographical structure of the county (i.e., being a rural or a densely populated county) are considered some of the main determinants of crime, thus we control for these variables in our crime regressions. In addition, we add the percentage of small firms on the total number of firms (by county),[43] where a small firm has less than 500 employees. Gruber (2001) [Section 2.4] reports that big firms have economies of scale for employer-based insurance, thus the employer-based insurance offering rate is higher in the case of big firms [Gruber (2001), Section 2.4 and Table 3]. Finally, to control for the variation in drug penetration and in risky behavior which might lead to higher crime rates [see Vitaro et al. (2001)], we control for: 1) the (per-capita) number of arrests for sale and possession of drugs[44] and 2) the (per-capita) number of arrests for gambling.[45] To adjust further for unobservable (time-invariant) variables, in our crime regressions we condition on county fixed effects, and we condition on year fixed effects to take into account crime trends.

[Table 3 about here]

insurance can be either Medicaid, CHIP (i.e., Children's Health Insurance Program), Tricare (for military personnel).

[36]The FBI uses the "hierarchy rule" to report and classify crimes: if an individual commits simultaneously two offenses (e.g., rape and aggravated assault (of another person)), the crime will be classified only as the most serious crime (i.e., rape). The most serious crimes (in decreasing order) are: 1) murder, 2) rape, 3) robbery, 4) aggravated assault, 5) burglary, 6) larceny, 7) car theft, 8) arson. The offenses of murder, car theft, and arson are exceptions to the Hierarchy Rule (i.e., they are reported separately). See http://www2.fbi.gov/ucr/ucr_general.html

[37]The data is available at http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s

[38]The data is available at http://censtats.census.gov/usa/usa.shtml (Accessed date: June-July 2014).

[39]The data is available at http://goo.gl/xwn8yo (Accessed date: June-July 2014).

[40]The percentage of people below the poverty line.

[41]The interactions of Hispanic-non Hispanic origin with White-Black-Asian-other race. Other race includes American Indian, Alaska Native, Native Pacific Islander, two or more races.

[42]Lee and Mccrary (2005) reports that individuals between 17 and 19 years old have the highest probability of committing crime; on the contrary, individuals older than 64 have the lowest probability of committing crimes.

[43]The raw data is available at www.sba.gov/sites/default/files/static_county.xls

[44]I.e., opium/cocaine, marijuana, synthetic drugs and other dangerous non narcotics.

[45]I.e., bookmaking horse and sports, numbers and lotteries and all other illegal gambling.

4 Methodology

The Fixed Effect Model
The linear unobserved effects model for T time periods take the form in (8):

$$y_{it} = \alpha_i + x_{it}\beta + \epsilon_{it}, \qquad i = 1, ..., N \ , \ t = 1, ..., T, \tag{8}$$

where i and t indicate the county and the year, respectively, and where β is a K-dimensional column vector; the K-dimensional row vector x_{it} contains the insurance coverage for county i at time t, i.e. $Coverage_{it}$, the controls (described below) and the time dummies. The outcome of interest, i.e. y_{it}, is the number of crimes per 1000 individuals, by county per year.

The WG (Within Group) Estimator
We assume the strict exogeneity of the explanatory variables conditional on α_i:

$$\text{Assumption WG.1: } E\left(\epsilon_{it}|x_i, \alpha_i\right) = 0, \quad t = 1, 2, ..., T, \quad \text{where} \quad x_i = (x_{i1}, x_{i2}, ..., x_{iT}). \tag{9}$$

in order to ensure that the WG estimator is well behaved asymptotically, we need a rank condition on the matrix of time-demeaned explanatory variables:

$$\text{Assumption WG.2: } \operatorname{rank}\left(\sum_{t=1}^{T} E\left(\ddot{x}_{it}'\ddot{x}_{it}\right)\right) = K. \tag{10}$$

where $\ddot{x}_{it} = x_{it} - \overline{x}_i$ is a K-dimensional row vector. The next assumption ensures that the WG estimator is efficient:

$$\text{Assumption WG.3: } E\left(\epsilon_i\epsilon_i'|x_i, \alpha_i\right) = \sigma_\epsilon^2 I_T \quad \text{with} \quad \epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, ..., \epsilon_{iT}), x_i = (x_{i1}, x_{i2}, ..., x_{iT}), \tag{11}$$

where I_T is the $T \times T$ identity matrix.[46]

The FE-IV Estimator
We assume the strict exogeneity of the instruments conditional on α_i:

$$\text{Assumption FE-IV.1: } E\left(\epsilon_{it}|z_i, \alpha_i\right) = 0, \quad t = 1, 2, ..., T, \quad \text{where} \quad z_i = (z_{i1}, z_{i2}, ..., z_{iT}). \tag{12}$$

The following rank conditions must be satisfied:

$$\text{Assumption FE-IV.2: } (a) \operatorname{rank}\left(\sum_{t=1}^{T} E\left(\ddot{z}_{it}'\ddot{z}_{it}\right)\right) = L; (b) \operatorname{rank}\left(\sum_{t=1}^{T} E\left(\ddot{z}_{it}'\ddot{x}_{it}\right)\right) = K. \tag{13}$$

where $\ddot{x}_{it} = x_{it} - \overline{x}_i$ is a K-dimensional row vector, and $\ddot{z}_{it} = z_{it} - \overline{z}_i$ is an L-dimensional row vector with $L \geq K$. Finally we need that

$$\text{Assumption FE-IV.3: } E\left(\epsilon_i\epsilon_i'|z_i, \alpha_i\right) = \sigma_\epsilon^2 I_T \quad \text{with} \quad \epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, ..., \epsilon_{iT}), z_i = (z_{i1}, z_{i2}, ..., z_{iT}), \tag{14}$$

where I_T is the $T \times T$ identity matrix.[47]

[46] For our empirical estimations, we cluster the standard errors at the county level in order to allow intra-cluster correlation of the error term, thus we relax Assumption FE.3.

[47] For our empirical estimations, we cluster the standard errors at the county level in order to allow intra-cluster correlation of the error term, thus we relax Assumption FE-IV.3.

The FD-IV Estimator
The vector of instrumental variables z_{it} must be uncorrelated with the error term in differences [Wooldridge (2010), p. 362]:

$$\text{Assumption FD-IV.1: For } t = 2, ..., T, \quad E(z_{it}'\Delta\epsilon_{it}) = 0 \tag{15}$$

where $\Delta\epsilon_{it} = \epsilon_{it} - \epsilon_{it-1}$. The following rank conditions must be satisfied in order that the FD-IV estimator is well behaved asymptotically:

$$\text{Assumption FD-IV.2: } (a)\,\text{rank}\left(\sum_{t=2}^{T} E\left(z_{it}'z_{it}\right)\right) = L; (b)\,\text{rank}\left(\sum_{t=2}^{T} E\left(z_{it}'\Delta x_{it}\right)\right) = K \tag{16}$$

where $\Delta x_{it} = x_{it} - x_{it-1}$ is a K-dimensional row vector , and z_{it} is an L-dimensional row vector, with $L \geq K$. Finally,

$$\text{Assumption FD-IV.3: } E\left(\epsilon_i\epsilon_i'|z_i, \alpha_i\right) = \sigma_\epsilon^2 I_{T-1} \quad \text{where } \epsilon_i \text{ is the } (T-1)\times 1 \text{ vector containing } \epsilon_{it}, t = 2, ..., T \tag{17}$$

an where I_{T-1} is the $(T-1) \times (T-1)$ identity matrix.[48]

The Two-Step Difference GMM Estimator
Starting form the panel data model (8) in first differences, we can construct a more efficient estimator than the FD-IV estimator.[49] The Difference GMM estimator takes the form in eq. (18) [Hayashi (2000)]:

$$\widehat{\beta}_{D-GMM} = \left(S_{XZ}'\widehat{W}^{opt}S_{XZ}\right)^{-1}\left(S_{XZ}'\widehat{W^{opt}}S_{XY}\right) \tag{18}$$

where

$$S_{XZ} = \frac{1}{N}\frac{1}{T-1}\sum_{i=1}^{N}\sum_{t=2}^{T} z_{it}'\Delta x_{it}; \qquad S_{XY} = \frac{1}{N}\frac{1}{T-1}\sum_{i=1}^{N}\sum_{t=2}^{T} z_{it}'\Delta y_{it}; \qquad \widehat{W}^{opt} = \left(\widehat{S}\right)^{-1} \tag{19}$$

and

$$\widehat{S} = \frac{1}{N}\frac{1}{T-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\widehat{\Delta\epsilon_{it}}\right)^2 z_{it}'z_{it} \tag{20}$$

Using the Difference GMM estimator we can improve with respect to the efficiency of the FD-IV estimator. The vector of instrumental variables must be contemporaneously exogenous in the first difference equation:

$$\text{Assumption GMM.1: For } t = 2, ..., T, \quad E(z_{it}'\Delta\epsilon_{it}) = 0 \tag{21}$$

We allow the dimension of the instruments to grow as t increases. The instrument matrix Z_i is given by

$$Z_i = \begin{pmatrix} z_{i2} & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & z_{i3} & z_{i2} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & z_{i4} & z_{i3} & z_{i2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad \text{or, collapsed,} \quad Z_i = \begin{pmatrix} z_{i2} & 0 & 0 & \cdots \\ z_{i3} & z_{i2} & 0 & \cdots \\ z_{i4} & z_{i3} & z_{i2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{22}$$

[48] For our empirical estimations, we cluster the standard errors at the county level in order to allow intra-cluster correlation of the error term, thus we relax Assumption FD-IV.3.

[49] The FD-IV estimator presented above is a special case of the GMM estimator, and it is obtained in the case that the optimal weighting matrix (to be presented below) is equal to the identity matrix: $\widehat{W}^{opt} = I_L$, where I_L is the $(L \times L)$ identity matrix.

In any case, the rank condition for the difference GMM estimator is [Wooldridge (2010), pp. 362]:

$$\text{Assumption GMM.2: } (a)\,\text{rank}\, E\left(Z_i' Z_i\right) = L; (b)\,\text{rank}\, E\left(Z_i' \Delta X_i\right) = K, \tag{23}$$

where ΔX_i is the matrix notation for Δx_{it}. We use the Windmeijer (2005)'s finite-sample correction for the two-step covariance matrix.

5 Health Insurance Coverage and Crime

In this section we use several estimators -in particular several Instrumental Variables estimators- which are suitable for panel data. Each of the IV estimators has some advantage over the other estimators: using the FE-IV estimator we do not lose the first period of our data to estimate the parameters of our model while the FD-IV and the Difference GMM estimators are more robust if the error term is serially correlated.[50] For reference, we also present the estimates of the WG (Within Group) estimator. Following Murray (2006) we use instruments that are strongly correlated with the insurance coverage and that come from various fields: some of the instruments come from the field of labor economics (e.g., self-employment), some instruments come from health economics (e.g., obesity at the county level). The use of instruments with different rationales tend to improve the robustness of the results [Murray (2006)].

5A Insurance Coverage and Crime: the WG (Within Group) Estimates

Table 4 shows the evolution of the insurance coefficient (β_0) for both the pooled measure of property crimes (obtained by summing up larceny, motor-vehicle theft, burglary, arson) and the pooled measure of violent crimes (obtained by summing up murder, aggravated assault, rape, robbery). Once we condition on the county FE, the insurance coverage has no effect on (pooled) violent crimes while the effect of the insurance coverage on (pooled) property crimes becomes stronger, the more detailed our model becomes.

[Table 4 about here]

The Altonji et al. (2005)'s selection on observables (method) for omitted variable bias

Altonji et al. (2005) and Bellows et al. (2008) derive a parametric representation of the omitted variable bias in the context of a non-linear model and in the context of a linear model, respectively. The idea is to confront the estimated coefficients of our main explanatory variable (i.e., insurance coverage) of two regressions: 1) the regression of the dependent variable (y) on our main explanatory variable (x) without any control, 2) the regression of (y) on our main explanatory variable (x) and on a set of observable controls (w). If the inclusion of the controls has no (or a very small) effect on the parameter estimate of x, then it is unlikely that omitted variables are driving the parameter estimate of x. Concisely, the higher is the *absolute value* of the ratio

$$\frac{\widehat{\alpha}_{OLS,C}}{\widehat{\alpha}_{OLS,NC} - \widehat{\alpha}_{OLS,C}}, \tag{24}$$

where $\widehat{\alpha}_{OLS,C}$ is the estimate of our main explanatory variable (x) in the case that we use additional controls and $\widehat{\alpha}_{OLS,NC}$ is the estimate of our main explanatory variable (x) in the case that we do not use any other control, the lower is the probability that our estimate of α is driven by omitted variable bias. The full explanation of the Altonji et al. (2005)'s method is reported in Appendix D.

The Omitted Variable Bias for Pooled Crimes: the results of the Altonji et al. (2005) technique

The Altonji ratio is equal to -3.440 in the case of the (pooled) property crimes and is equal to -4.567 in the

[50] Additionally, we do not need that the instruments are strictly exogenous, the instrument only needs to be contemporaneously uncorrelated with the error term in differences.

case of the (pooled) violent crimes[51]. Since the absolute value of the ratio is higher than 1, we are confident that the results are not driven by omitted variable bias for our aggregate measures of crime.

Estimates by crime type

The results of the WG estimator using time dummies and the full set of controls are reported in Table 5 for each type of crime[52]. The insurance coverage has a positive impact on the number of larcenies and car thefts while it has a negative impact on the number of robberies. Individuals/families that become insured have a lower disposable income (after the insurance gain), which might incentivize them to commit pecuniary crimes such as larcenies and car thefts. Finally, burglary is the most dangerous property crime since the burglar risks to be killed while stealing in a house, thus individuals might be disincentivized to commit burglary.

The increase of the insurance coverage in a county leads to better health in the community, which in turn might increase the social capital of that community [Carpiano (2007), pp. 650-651]; in addition, there is some evidence that higher social capital in a community leads to a lower level of violent crimes [Sampson et al. (1997)]. An increase in the insurance coverage leads to a lower amount of robberies while it has no effect on rapes, murders and aggravated assaults. Most of the murders and aggravated assaults are drug-related [Goldstein (1998)][53], thus we find counterintuitive that changes in insurance coverage have no effect on the number of murders and aggravated assaults.

[Table 5 about here]

Omitted Variable Bias for Each Crime Type: the Altonji et al. (2005) technique

The absolute value of the Altonji ratio is lower than 1 in the cases of burglary, rape and robbery,[54] thus omitted variable bias might be present for these types of crime. For the remaining five types of crime, the Altonji ratios range from -1.545 (arson) to -16.207 (aggravated assault), thus it is unlikely that omitted variable bias is driving the results for these crime types.

Measurement Error Bias and Reverse Causality Bias

The presence of measurement error in an explanatory variable causes an attenuation bias, thus the *true coefficient*[55] of the insurance coverage in the cases of larceny, car theft and robbery would be even higher (in absolute value) than in Table 5 if $coverage_{it}$ is measured with error. Thus, the presence of measurement error would not drastically change the interpretation of our results in the cases of larceny, car theft and robbery.[56]

If there is reverse causality between crime and the insurance coverage, we can rewrite our model as a system of equations:

$$\ddot{Crime}_{it} = \ddot{Coverage}_{it}\beta_0 + \ddot{\epsilon}_{it} \tag{25}$$

$$\ddot{Coverage}_{it} = \ddot{Crime}_{it}\beta_1 + \ddot{v}_{it} \tag{26}$$

where, for a generic random variable x_{it}, we have that $\ddot{x}_{it}$ is represented by $\ddot{x}_{it} = x_{it} - \overline{x}_i$. If β_1 is significantly different from zero, our WG estimate of β_0 will be inconsistent[57]. We hypothesize that β_1 *is negative in the cases of larceny theft, car theft, burglary and robbery* since an increase in crime incentivizes firms and middle-high income individuals/families to relocate to safer counties, which in turn will decrease the percentage of individuals covered by insurance. If our hypothesis is confirmed (by the data), the WG estimates of β_0 are

[51]As in Nunn et al. (2012) [Table 4], the county FE and the time dummies are used in both the regression with controls (where we have β^C) and without controls (where we have β^{NC}). In the case of property crimes, the Altonji ratio is found using the coefficient in Column (4) of Table 4 as our $\hat{\beta}^C$, and using the coefficient in Column (3) of Table 4 as our $\hat{\beta}^{NC}$, and using the ratio $\frac{\hat{\beta}^C}{\hat{\beta}^{NC}-\hat{\beta}^C}$.

[52]Using the Hausman test, we checked whether the WG estimator or the Random Effects estimator should be used for each type of crime; the WG estimator must be used for all crime types. The results are available from the author.

[53]Goldstein (1998) [page 1926]: "About one-half of all violence in all studies was drug-related" and Goldstein (1998) [page 1919]: "Most systemic violence arises from doing business in a black market. Examples of systemic violence include territorial disputes between rival dealers, assaults and homicides committed within particular drug-dealing operations in order to enforce normative codes, robberies of drug dealers, punishment for selling adulterated or bogus drugs, assaults to collect drug-related debts, and so on."

[54]The Altonji ratios are equal to: a) 0.481 (burglary), b) -0.390 (rape), c) -0.521 (robbery).

[55]We define the true coefficient as the coefficient that is obtained through an IV estimation, where the instruments are strong and exogenous.

[56]Anyhow, it might change the interpretation of the insurance coefficient (β_0) in the case of the remaining crime types.

[57][...] since $E\left(\ddot{Coverage}_{it}\ddot{\epsilon}_{it}\right) = E\left(\left(\ddot{Crime}_{it}\beta_1 + \ddot{v}_{it}\right)\ddot{\epsilon}_{it}\right) = \beta_1 E\left(\ddot{Crime}_{it}\ddot{\epsilon}_{it}\right) + E\left(\ddot{v}_{it}\ddot{\epsilon}_{it}\right) \neq 0$ even in the case that $E\left(\ddot{v}_{it}\ddot{\epsilon}_{it}\right) = 0$.

downward biased.[58] We abstain from formulating hypotheses in *the case of arson, murder, rape and aggravated assault* since the subjects of rapes, arsons and aggravated assaults might develop psychological and emotional problems which lead to a higher consumption of health care services[59] and potentially to a higher probability of buying health insurance. The latter effect contrasts with the previous effect (i.e., the drop in coverage due to the relocation of firms/individuals), thus we do not know whether the WG estimate of β_0 suffers from downward or upward bias in the cases of arson, murder, rape and aggravated assault. Finally, highly likely the WG estimates of the insurance coefficient (β_0) suffer from omitted variable bias in the cases of burglary, rape and robbery [see the Altonji ratios described above]; thus, our hypothesis of a downward biased WG estimate of β_0 might not show up in the estimates because of the additional omitted variable bias in the cases of burglary, rape and robbery.[60]

5B Health Insurance and Crime: the FE-IV estimates

To identify the impact of the insurance coverage on crime using instrumental variables, one needs a set of instruments that predicts changes in insurance coverage, but is unrelated to changes in crime after controlling for confounders and unobserved (time-invariant) heterogeneity. First, we use a restricted set of instruments composed by only one instrument, which is not subject to the weak instruments critique. Secondly, we introduce other strong instruments to improve the first-stage fit of the data and to show that our results are consistent using instruments with different rationales á la Murray (2006).

The FE-IV Estimates Using One Instrument

The self employment status is correlated with the insurance coverage rate since the insurance expenditure's tax treatment differs between employees and self employed individuals.[61] Self employed individuals operate in three organizational forms -sole proprietorships, partnerships, and corporations- and the probability of being insured depends on the organizational form in which the self employed individual operates [Meer and Rosen (2004)]. Self employed individuals that operate as a corporation have a high probability of being covered by health insurance since the insurance expenditure is fully deductible [Meer and Rosen (2004)]. On the contrary, self employed individuals that operate as a member of a partnership or as a sole proprietor might have a lower probability of buying health insurance (than employees) since the insurance expenditure is not fully deductible[Meer and Rosen (2004)].[62] We find that in counties with high percentages of self employed individuals, the insurance coverage is higher than in counties with fewer self employed individuals (i.e., the instrument is strong) [see Table F3].

Non-agricultural self employment has been rising since the early 1970s [Blau (1987), Figure 1];[63] the three main causes of this trend are changes in the industrial structure, technological change and changes in marginal tax rates [Blau (1987)]. Changes in the industrial structure have favored industries in which scale economies are relatively unimportant and small firms are viable [Blau (1987)].[64] Technological change (i.e., the introduction of personal computers) increased the competitiveness of small firms in many industries [Blau (1987)].[65]

[58] Confronting the results of the Within Group estimator with the results of the FE-IV estimator, there is evidence of downward bias in the estimate of β_0 in the cases of larceny, car theft and robbery.

[59] [...] and also the relatives of murdered individuals might develop psychological and emotional problems which lead to a higher consumption of health care services.

[60] The idea is to compare the WG estimates and the corresponding IV estimates to understand whether there is downward or upward bias of the WG estimator; anyhow, the difference between the WG and the IV estimates might also depend on omitted variable bias in the cases of burglary, rape and robbery, thus in these cases it is more difficult to draw a conclusion on reverse causality bias.

[61] The data on self employment is takes from the Bureau of Economic Analysis [see `http://goo.gl/ewLTBU`].

[62] The insurance expenditure is fully deductible in the case of employees.

[63] This trend continued also after the publication of Blau (1987), and up to the present day. Using our data, we obtain that the self employment rate was equal to 21.50% in 2005 and it grew to 24.21% in 2010 (with a positive growth in each panel year), and it grew even further after 2010 reaching a self employment rate of 24.9% in 2013.

[64] In addition, Blau (1987): "increased wage rigidity has increased the proportion of the labor force that resorts to self-employment as a response to being rationed out of wage jobs."

[65] The subsequent introduction of the internet might have even increased this trend.

Higher marginal tax rates increases the attractiveness of self employment since an increase in marginal tax rates incentivizes individuals to under-report income, and it is easier to under-report income for self employed individuals than for employees [Blau (1987) p.457].[66]

The main determinants of crime are the size of the police force, the incarceration rate, drug shocks [Levitt (2004)]. We hardly see any relationship between the three determinants of self employment and the determinants of the crime rates. We do not see any relationship between technological change (and the industrial change) on one side and the determinants of crime rates while there might be a relationship between taxation changes and other government policy changes (e.g., an increase in the size of the police forces). In the case of taxation, if a newly elected government changes the size of the police forces in period t, and the government changes the marginal tax rates in period t (which impacts on the probability of becoming self employed), then the instrument $Self\ Employment_{it}$ might be endogenous. Anyhow, we must underline that the size of the police forces are decided by the federal and the local governments [Levitt (2002)] while the marginal tax rates are decided by the federal government, decreasing the probability (and the strength) of endogeneity of $Self\ Employment_{it}$.[67] In order to break the correlation between the government changes of the police forces and the changes in marginal tax rates, we use $Self\ Employment_{it+2}$ rather than self employment in period t as our main instrument. We report the reduced-form estimates in Table 6 and the second-stage estimates in Table 7.

[Table 7 about here]

Reduced-Form and Second-Stage IV estimates

The first two columns of Table 6 show that $Self\ Employment_{it+2}$ is positively correlated with the larceny and the car theft rates while $Self\ Employment_{it+2}$ is unrelated with the remaining crimes. If we assume that the exclusion of $Self\ Employment_{it+2}$ from the crime equations is valid, the combination of a positive first-stage coefficient and a positive reduced-form coefficient implies that the IV second-stage estimate of coverage will be positive.

An increase of the insurance coverage increases the larceny theft and the car theft rates, and the impact of the insurance coverage on these two crime rates is much stronger than in the case of the WG estimator [see Table 5 and Table 7]; thus, our hypothesis of reverse causality bias in the cases of larceny theft and car theft is confirmed by the data.[68] Changes in insurance coverage have no effect on the remaining crime rates. To conclude, once we take the endogeneity of the insurance coverage into account, we find that an increase of the insurance coverage increases the number of property crimes[69], and in particular it strongly increases the number of the less risky property crimes (e.g., car theft and larceny theft) while it has no effect on more risky property crimes (i.e., burglary and arson) and on violent crimes.

A Further Check

Given the previous discussion, we think that our initial instrument is a valid instrument; to give further validation on the validity of our main instrument (i.e., $Self\ Employment_{it+2}$), we report the FE-IV estimates where we use the female and the male obesity at the county level as our instruments, and we compare the second stage results with Table 7. The obesity instruments are strong [Table F4] -even though each of the two instruments is weaker than $Self\ Employment_{it+2}$- and the over-identifying restriction is valid [see the Hansen test in Table F5], thus it seems that the obesity instruments are valid.[70] The second-stage results using the two obesity instruments point in the same direction as the results of Table 7 (i.e., an increase of the insurance coverage implies an increase of larcenies and car thefts only), giving further validation to the validity of our

[66] The higher are the marginal tax rates, the higher is the share of self employment in the U.S. [see Table 2 of Blau (1987)].

[67] In addition, if the government changes both the marginal tax rates and hires more police in period t, the drop in crime might be immediate (i.e., deterrence effect of the police) while the change in the self employment rate will be more gradual (i.e., individuals need months/years and they need capital before they set up their own business), decreasing even further the probability of endogeneity of $Self\ Employment_{it}$.

[68] In fact, the insurance coverage is endogenous at the 10% significance level in the case of larceny and car theft [see the Hausman test in Table 7].

[69] [...] because of the disposable income drop after the insurance gain.

[70] A more detailed discussion on the exogeneity of the obesity instruments is presented below.

main instrument (i.e., $Self\ Employment_{it+2}$) [Murray (2006)].

The FE-IV Estimates Using the Self Employment and Diabetes Instruments

We add another instrument to form a set of two instruments so that we can test the over-identifying restriction. The additional instrument is the diabetes incidence rate in period $t-1$ -i.e., the new cases of diabetes relative to the total number of individuals living in a county- which is a strong instrument [Table F6],[71] and which might satisfy the exogeneity condition.[72] Diabetes is mainly caused by genetics, viruses and toxins, and by lack of physical activity [American Diabetes Association (2008); we do not see any reason why the set of factors that influences the diabetes incidence (rate) would also affect crime, especially after we control for unobserved heterogeneity and for several socio-economic confounders. In addition, we use the lagged diabetes incidence rate (i.e., $Diabetes\ Incidence_{it-1}$) to make our instrument less susceptible to endogeneity concerns. The results of the Hansen J tests reported in Table 8 seem to confirm that our instruments are exogenous since all the p-values of the Hansen J statistics are larger than 0.10. We report the second-stage estimates in Table 8 and the reduced-form estimates in Table F7.

[Table 8 about here]

Reduced-Form and Second-Stage IV estimates

Using the reduced-form estimates, $Self\ Employment_{it+2}$ is still significant in the case of the larceny and the car theft rates even after we add $Diabetes\ Incidence_{it-1}$ to our set of instruments.[73]

The sign of the second-stage coefficients for the coverage rate are the same in Table 7 and Table 8, and also the magnitude of the coefficients are close between Table 7 and Table 8, which reinforces the validity of our instruments [Murray (2006)]. An increase of the insurance coverage implies an increase of the larceny and car theft rates while changes of the coverage rate do not have any effect on the other crime rates. The insurance coverage is endogenous in the cases of larceny theft, car theft and rape. Finally, once we add our second instrument, the standard errors are substantially reduced.

The FE-IV Estimates Using Five Instruments

We add ***a)*** the percentage of obese females in period t, ***b)*** the percentage of obese males in period t, ***c)*** the percentage of obese females in period $t-3$ to our instrument set.[74] Obese individuals might be less concerned about their health than non-obese individuals, so that the low interest in their own health is related to their obesity. In this case, counties with high obesity rates would be characterized by low (health insurance) coverage rates since obese individuals are not concerned about their health. On the other hand, obese individuals might experience severe problems in daily living activities which might incentivize them to consume more health care and to buy health insurance. In this case, counties with high obesity rates would be characterized by high coverage rates. Our hypotheses seem to be supported by the data since the obesity instruments are strongly (positively and negatively) correlated with the insurance coverage [see the F-test in Table F8].

There are two main determinants of obesity: a) an excessive caloric intake, b) an inadequate amount of physical activity;[75] both determinants are in turn affected by various factors that we summarize in Appendix G. We do not see any relationship between the determinants of obesity (see Appendix G) and the determinants of crime (reported above), thus we believe that the obesity instruments are exogenous. In addition, we control for many socio-economic confounders such as the county poverty rate and the county personal incomes, which

[71] In addition, the set of instruments is strong since the F-test of excluded instruments is substantially higher than 10.

[72] The data for the diabetes incidence is available at `http://www.cdc.gov/diabetes/atlas/countydata/County_ListofIndicators.html`. We use the estimates of the age-adjusted diabetes incidence rate.

[73] $Diabetes\ Incidence_{it-1}$ is not significant in the reduced-form in the case of the larceny and the car theft rates, anyhow this is caused by the fact that the first-stage coefficient of $Diabetes\ Incidence_{it-1}$ is very close to zero (but significant) [see Table F6].

[74] The data for female and male obesity at the county level is taken from the "Additional file 4" of the paper by Dwyer-Lindgren et al. (2013), which is available at `http://www.pophealthmetrics.com/content/11/1/7/additional`.

[75] See `http://frac.org/initiatives/hunger-and-obesity/what-factors-contribute-to-overweight-and-obesity/` for a detailed description of the causes of obesity.

decreases further the probability of endogeneity of the obesity instruments.[76] The exogeneity of the obesity instruments seems to be confirmed by the Hansen J tests reported in Table 9: the over-identifying restriction is valid at the 10% level for all crimes, and the p-values of the Hansen J tests are high. We report the second-stage estimates in Table 9 and the reduced-form estimates in Table F9.

[Table 9 about here]

Second-stage IV estimates

The sign of the insurance coefficient (for all crime types) is the same in Table 8 and Table 9, thus reinforcing the validity of our instruments [Murray (2006)]. Our main results are confirmed once we add the obesity instruments to our previous instrument set: an increase of the insurance coverage implies an increase of the larceny and car theft rates while it has no effect on the other crime rates. The magnitude of the coefficients are different with respect to Table 8, anyhow the difference is not very large for all crime types. Finally, the insurance coverage is endogenous at the 5% significance level in the cases of larceny and car theft.

5C The FD-IV and the Difference GMM estimates

The results of the FD estimator are reported in Table F10 as a reference for the FD-IV and the GMM estimators.

The FD-IV estimates

In this subsection we check the consistency of the results of Section 5B using the FD-IV estimator. The First Difference model is suitable for panel data since it deletes unobserved heterogeneity from the model, thus it decreases the probability that the explanatory variables (in differences) are correlated with the error term (in differences). The FD-IV estimator has one advantage with respect to the FE-IV estimator: the instrument z_{it} needs to be uncorrelated with respect to $\Delta\epsilon_{it} = \epsilon_{it} - \epsilon_{it-1}$ rather than with all the error terms (i.e., $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, ..., \epsilon_{iT})$; see eq. (12)).[77]

We use four instruments:[78] 1) $\Delta Hospital\ Days\ Retarded_{it}$ (i.e., the differenced data for the number of hospital days for individuals with mental retardation), 2) $\Delta Home\ health\ agencies_{it}$ (i.e., the differenced data for the number of home health agencies), 3) $\Delta Female\ obesity_{it}$ (i.e., the differenced data for the percentage of female obesity in each county), 4) $Self\ employment_{it}$. In the previous discussion we discussed the rationale for the exogeneity of instrument 3) and instrument 4); the number of home health agencies is correlated with the insurance coverage since the insurance gain implies an increased consumption of hospital care which is a substitute of home health agencies' care. On the other hand there is little reason to think that changes in crime and changes in the number of home health agencies are directly correlated once we control for time trends in crime, unobserved heterogeneity, and for several socio-economic confounders. The number of hospital days for mentally retarded individuals is correlated with the insurance coverage: if the number of mentally retarded individuals -proxied by the number of hospital days for mentally retarded individuals- increases in a county, the insurance coverage will increase since these subject benefit the most from having health insurance. The causes of mental retardation are mainly genetic abnormalities, such as the Down syndrome, or being subjected to a prenatal trauma, thus we do not see a direct relationship with the amount of property/violent crimes. In old crime literature, mental retardation was associated with a high probability of committing arson; anyhow, in recent crime literature, the association between intellligence and the probability of committing arson is discarded [Rasanen et al. (2010)].[79] The second-stage results are reported in Table 10 and the first-stage IV

[76] Counties with high obesity rates are characterized by high poverty rates [Levine (2011)]; in addition, female obesity rates are not only correlated with poverty rates, but are more generally correlated with incomes [Jeffrey and French (1996)]. If we do not control for poverty rates and incomes, and since poverty rates and incomes are determinants of crime [see Section 2A], then our instruments might be invalid.

[77] On the other hand, we lose one time period in the case of the FD-IV estimator, and the FE-IV estimator is more efficient if the assumptions in eq. (12)-(13)-(17) are satisfied (and if $t > 2$).

[78] We do not use the set of three instruments used in Section 5B since these instruments are not strong in the case of the FD-IV estimator. The last findings do not imply that the instruments used in Section 5B are not valid in the case of the FE-IV estimator since we are using a different number of years (5 versus 6 years) and different assumptions between the FD-IV and FE-IV estimators.

[79] Rasanen et al. (2010) [pp. 619]: "We speculate that currently forensic psychiatric examinations are carried out more frequently than previously when only severely mentally disordered individuals and those with very low IQs were examined. So the examination material has become healthier

estimates are reported in Table F11. The instruments are strong as shown in Table F11.

[Table 10 about here]

The results of Table 10 confirm our main findings reported in Section 5B: in the cases of larceny and car theft, the coefficient of the insurance coverage is positive and significant at the 5% level, and the magnitude of the coefficient of the insurance coverage in the cases of larceny and car theft are very close between the FE-IV estimator [see Table **??**] and the FD-IV estimator [see Table 10].

The Difference GMM Estimates

In the previous discussion we used the FD-IV estimator with instruments constructed using random variables in periods t and $t-1$, thus we assumed that the instruments were strictly exogenous. We can use milder assumptions for the instruments since we only need that $E(z_{it}\Delta\epsilon_{it}) = 0$ when we use our model [eq. (8)] in first differences, for example we can use instruments constructed using random variables in period $t-2$ or in period $t+1$.[80] In the following discussion we do not use instruments from periods t and $t-1$ to increase the probability that our instruments are exogenous. In addition, we use both *internal instruments* such as $\Delta Coverage_{it-3}$ and *external instruments* in order to improve the strength of the instrument set.[81] We do not use the (static) System GMM since the System GMM introduces an additional assumption, $E(\alpha_i z_{it}) = 0$, that highly likely is not satisfied.[82]

Using the Difference GMM estimator we estimate the following equation: $\Delta Crime_{it} = \Delta Coverage_{it}\beta_0 + \Delta x_{it}\beta_1 + \Delta\epsilon_{it}$, and we instrument $\Delta Coverage_{it}$ with (collapsed) instruments to be specified below.[83] When the error term in levels (i.e., ϵ_{it}) is not autocorrelated or it is autocorrelated of order one (i.e., larceny theft, car theft, burglary, arson and murder), we use the following mix of internal and external (collapsed) instruments: 1) $\Delta Coverage_{it-3}$, 2) $Coverage_{it+2}$, 3) $Coverage_{it+3}$, 4) $Self\ Employment_{it-2}$, 5) $Self\ Employment_{it-3}$, 6) $\Delta Retardation_{it-3}$. When the error term in levels is autocorrelated of order two (i.e., rape, aggravated assaults and robbery), it is better not to use internal instruments from period $t-3$ up to period $t+2$, thus we change our set of instruments to the following one:[84] 1) $Coverage_{it+3}$, 2) $Coverage_{it+4}$, 3) $\Delta Self\ Employment_{it-3}$, 4) $\Delta Self\ Employment_{it-4}$, 5) $\Delta Retardation_{it-3}$, 6) $\Delta Retardation_{it-4}$, 7) $\Delta Female\ obesity_{it-3}$, 8) $\Delta Female\ obesity_{it-4}$. Finally, we report both the results of the one-step and two-step GMM estimators in Table 11:[85] the results of the one-step GMM estimator are similar to the results of the FD-IV estimator [Table 10] which reinforces the robustness of our IV results [Murray (2006)].[86]

[Table 11 about here]

The instrument set seems to be valid since the null hypothesis of the Hansen J test is never rejected at the 5% level for all crime types. Increases in the insurance coverage increase the number of larceny thefts, car thefts and robberies. The first two results confirm our previous findings while the latter result is a new finding. Robbery is categorized as a violent crime even though the main goal of robbery is not violence per se but a pecuniary gain, thus the positive and significant coefficient of coverage in the case of robbery does not contrast with our previous findings. Finally, the standard errors of the two-step GMM estimator are approximately 10% lower than the standard errors of the one-step GMM estimator, and the coefficients of the insurance coverage

over the years and arsonists nowadays are found to have more normal intelligence levels."

[80] The choice of the instruments depends also on the autocorrelation of the error term in levels (i.e., ϵ_{it}); more details are provided below.

[81] We rely on (internal or external) instruments from periods $t-2$ to $t-4$, and from periods $t+2$ to $t+3$. Additionally, in the case that the error term (in levels) is autocorrelated of order 1, then we do not use *internal instruments* in periods $t-2$, $t-1$, t and $t+1$. If the error term (in levels) is autocorrelated of order 2, then we do not use *internal instruments* in periods $t-3$ up to $t+2$, and so forth in the case that the error term (in levels) is autocorrelated of order 3. These exclusions are made to avoid endogeneity of the internal instruments.

[82] In fact, using the (static) System GMM estimator we find that the null hypothesis of the Hansen J test is rejected at the 5% level using many combinations of instruments.

[83] We collapse the instruments in order to limit the number of instruments. Instrument proliferation have negative effects on the power of the Hansen test; in addition, using a high number of instruments increases the probability that the set of instruments is invalid.

[84] The higher order of correlation of the the error term makes the internal instruments more suspect.

[85] The Windmeijer (2005)'s small sample correction for the covariance matrix is used in the case of the two-step GMM estimator.

[86] The fact that the estimates of the two-step GMM estimator are (quite) different from the estimates of the FD-IV estimator does not invalidate our findings since the one-step and the two-step GMM estimators coincide under very restrictive (and unrealistic) assumptions.

vary substantially across the two methods.

6 Theoretical and Empirical Evidence of Moral Hazard in the U.S. Health Insurance Market

6A Theoretical Evidence of Moral Hazard

A group of uninsured low-income adults in Oregon was selected by lottery in 2008 to be given the chance to apply for Medicaid [Finkelstein et al. (2012)]; this is the only natural experiment which shows the effects of becoming insured on health care utilization in the U.S.. Becoming insured increases the probability of hospital admissions by 30% [Finkelstein et al. (2012)] and it implies a 778$ increase in annual health care spending, that is, a 25% increase in annual health care spending relative to the control group [Finkelstein et al. (2012), page 23]. Also the drug utilization and the compliance with the preventive health tests increases substantially after the insurance gain. Using the 1994, 1996 and 1998 waves of the MEPS (Medical Expenditure Panel Survey) dataset and using IV techniques, Meer and Rosen (2004) find that the (public or private) insurance gain increases the probability of having a cholesterol check by 39.9% and it increases the probability of visiting an office-based care provider by 35.2% [Table 3 of Meer and Rosen (2004)]. Additionally, the insurance gain implies an increase of all types of preventive health tests. Using the 2000-2008 waves of the MEPS dataset, Jerant et al. (2013) find evidence of *ex-post moral hazard* (i.e. increased use of health care because of shared costs with the insurer) while the authors do not find evidence of *ex-ante moral hazard* (i.e. riskier behavior - such as starting to smoke- because of reduced personal costs from risky behavior).[87] For example, the insurance gain increases the probability of taking the mammography test by 10.4%.

Kreider et al. (2009) use the partial identification approach pioneered by Manski (1990) to study the effect of the expansion of health insurance coverage on health care utilization. If the universal health coverage is achieved in the U.S., the per-capita monthly provider visits would rise by 8 percent, and the mean health expenditures per month would rise by 16 percent *across the non-elderly population*.[88] Hadley et al. (2008) estimate how much the health care expenditure would increase in the U.S. if universal coverage is achieved.[89] The simulations of Hadley et al. (2008) suggest that, for *people who were uninsured and that they subsequently gained insurance*, the total spending per person would increase by 70% (i.e., from 2290$ to 3885$ per person per year) once they gain (public or private) health insurance [Exhibit 5 of Hadley et al. (2008)]. Other papers that show a positive (and significant) relationship between having (public or private) health insurance and health care expenditure are Dong (2000) and Ward et al. (2007).

To conclude, the (public or private) insurance gain implies a large increase in health care expenditure. The increased health care expenditure connected with the private insurance gain will be repaid by the insuree through the premium, and it will be repaid with higher per-capita taxes (or higher public debt) in the case that the insuree gains public health insurance. Using the macroeconomic theory of Barro (1974)[90], the higher per-capita taxes (or the higher public debt) will decrease the disposable income (for non health care goods) of the U.S. society. In turn, the decrease in disposable income might have an effect on crime [Section 2A]. Finally, the insurance gain has an effect on job mobility [see Appendix A].

6B Empirical Evidence of Moral Hazard in the U.S.

In this section we show the presence of (ex-post) moral hazard, that is, an increase of the insurance coverage implies an increase of health care utilization. Since we lack data on physicians' services, we focus on hospital

[87] The health care costs are shared between the insurer and the insuree through the co-payments (or the co-insurance rates).

[88] We must underline that these values are for the overall population, not for the individuals that gain insurance.

[89] Under the assumption that the currently uninsured are similar to the lower and lower-middle income insured individuals. In fact, most of the uninsured individuals have low-middle incomes [see Exhibit 4 in page 5 of Holahan (2010)].

[90] I.e., an increase in taxes (in this year) is equivalent to an increase in debt (in this year).

and nursing home utilization data from the AHRF (Area Health Resource File) dataset.[91] Most of the health care expenditure is devoted to hospital and nursing home utilization, not to physicians,[92], thus the dependent variables that we use in this section account for the majority of total health care utilization. Finally, the nursing home utilization is not limited to individuals that are older than 64; Kemper and Murtaugh (1991) shows that also individuals that are younger than 64 use nursing home services.[93] We describe the health utilization measures, that is, all the dependent variables used in this section, in Appendix E. We present the descriptive statistics for all the variables used in this section in Table F1.

Methodology

Many individuals move to other counties in order to utilize hospital services since 19.74% of the counties (i.e., 620 counties) are not served by any hospital during 2005-2010.[94] We do not have data on the number of nursing homes by county; anyhow, McAuley et al. (2002) [pp. 47 and Table 1] show that 35.3% of the rural counties in Virginia are not served by any nursing home. *Even if there is moral hazard*, the increase of the insurance coverage in a county which is not served by any hospital (or nursing home) will not increase the hospital (or nursing home) utilization measures simply because the county is not served by any hospital. Thus, we use a linear model which does not contain county fixed effect,[95] but we use state-time dummies to control for differential state trends in the utilization of health services and we use year dummies to control for time trends in the U.S.:[96]

$$Utilization_{it} = Coverage_{it}\beta_0 + x_{it}\beta_1 + \delta_t + \mu_{s,t} + v_{it}, \qquad i = 1, ..., N;\ t = 1, ...T \tag{27}$$

where, as in Meer and Rosen (2004), x_{it} contains the following controls: **a)** eight dummies for the races in each county where the baseline is White and Non-Hispanic, **b)** gender, **c)** four dummies for the age distribution in each county;[97] finally, we add the **d)** population density in order to control for differences in rural and urban health care utilization.

There are two possible reasons why the health care utilization impacts on the insurance coverage (i.e., reverse causality bias): 1) insurers might be able to identify individuals that will be heavy health care users, thus insurers might ask a very high premium to this group of individuals so that they will not buy insurance, 2) the anticipation of high health care utilization incentivizes the individual to buy insurance, leading to an upward bias in the estimate of the insurance coefficient (β_0). Also, we cannot exclude a priori the presence of omitted variable bias and measurement error. Therefore, we use the instrumental variables (IV) technique to estimate the effect of insurance coverage on utilization.

Meer and Rosen (2004) show that self-employment status has a strong impact on the probability of buying health insurance since employees have a more favorable tax treatment of health insurance costs than self-employed individuals;[98] in addition, Meer and Rosen (2004) show that health care utilization does not differ between employees and self-employed individuals, thus the self-employment status might be a valid instrument for the health insurance coverage. Based on the results of Meer and Rosen (2004) we use two

[91] Available at `http://ahrf.hrsa.gov/download.htm`.

[92] Anderson et al. (1999) [page 182]: "The United States spent 42.2% of its health care dollar on hospital services in 1996". Using data from the CMS (Centers for Medicare & Medicaid Services) we find that the expenditure in nursing home care is approximately 7.2% of the total health expenditure in the U.S. [see `http://goo.gl/LGEVGx`]

[93] Using the 1986 National Mortality Followback Survey, Kemper and Murtaugh (1991) [Table 1] finds that, among people that died before they reached 65 years old, 7% of individuals used nursing home services.

[94] In addition, approximately 47% of the counties contain only one hospital and approximately 16% of the counties contain two hospitals in each panel year. In the case that only one or two hospitals are available in a county, sick individuals might move to other counties to obtain hospital services in the case that they have uncommon diseases, which are treated only in few hospitals.

[95] We do not use county fixed effects since the within-variation of the insurance coverage is substantial while the within-variation of the utilization measures is zero for all the counties without hospitals (or nursing homes).

[96] Also, we use the full panel. If we use a restricted panel containing only the data for counties with one or more hospitals, the estimation would suffer from omitted variable bias since $Utilization_{it} = Coverage_{it}\beta_0 + x_{it}\beta_1 + \delta_t + \mu_{s,t} + (Coverage_{it}^{No\ hospital}\beta_2 + \eta_{it})$, where $(Coverage_{it}^{No\ hospital}\beta_2 + \eta_{it})$ is the new error term, and the coverage in counties with no hospital ($Coverage_{it}^{No\ hosp.}$) is correlated with $Coverage_{it}$ and with $Utilization_{it}$.

[97] The four dummies are the percentages of the: 1) population between 15 and 19 years old, 2) population between 20 and 24 years old, 3) population between 25 and 29 years old, 4) population older than 64 years old.

[98] I.e., the employer contributions for the employer-based health insurance is not taxed while there are no deductions for individual health insurance.

instrumental variables for the insurance coverage: 1) $Self\ Employment_{it-1}$, where self-employment is defined as the fraction of the number of non-farm proprietors on the total employment (at the county level)[99] 2) $Self\ Employment_{it}^{Neighbor}$, which is the fraction of self-employed individuals on the total employment in the neighboring county (in year t).[100] The instruments satisfy the relevance condition [see Table F2],[101] and the instruments might satisfy the exogeneity condition since the null hypothesis of the Hansen J test is not rejected at the 5% significance level [Table 12-13].[102]

Second-stage IV Estimates

The results of the second-stage regressions are reported in Table 12-13.

[Tables 12-13- about here]

The insurance coverage is endogenous at the 10% significance level in the case of all the health care utilization measures except in the case of the LT (long term) rehabilitation inpatient days [see the p-values of the Hausman tests in Table 12-13].

An increase of the insurance coverage implies an increase of all health care utilization measures (with the exception of the LT rehabilitation inpatient days), thus we show substantial evidence of (ex-post) moral hazard in the U.S.. Individuals which experienced traumas such as brain injury, spinal chord injury or a stroke, might use hospital LT rehabilitation services regardless of their health insurance status simply because these individuals might die in the absence of rehabilitation. The previous reasoning might explain why the insurance coverage has no effect on the number of LT (Long Term) rehabilitation inpatient days.[103]

7 Checking for behavioral changes caused by the insurance gain

In the previous discussion we focused on the effect that the insurance gain has on crime, and we hypothesized that the insurance gain affects crime through moral hazard, that is, the increased health care utilization - following the increase of coverage in a county- implies a lower disposable income which in turn implies a higher larceny and car theft rate. In this section we check whether the insurance gain implies another behavioral change: we check whether an increase of the insurance coverage increases the sloppiness of individuals. If individuals become less careful after the insurance gain, then there might be an increase in property crime rates; for example, individuals might not use some precautions that would inhibit criminals to steal (in) their property.[104]

After an individual becomes insured, she might be less careful about her health behaviors (i.e., ex-ante moral hazard), and she might be less careful about her overall behavior.[105] To test the latter behavioral effect, we use the (state-level) pedestrian fatality rate as our dependent variable (taken from the National Highway Traffic Safety Administration),[106] and we check whether changes in the insurance coverage have an effect on the pedestrian fatality rate. If individuals are characterized by a less careful behavior after the health insurance

[99] The data for the number of non-farm proprietors and for the total employment are available from the BEA (Bureau of Economic Analysis) at `http://goo.gl/kMlsBC`.

[100] The AHRF dataset includes the identifier codes of the neighboring counties for each of the 3141 counties. A county can have a minimum of one neighboring county up to a maximum of 14 neighboring counties. For simplicity we use only the self-employment of the 1^{st} neighboring county, that is the county that is categorized as the 1^{st} neighbor in the AHRF dataset.

[101] The statistic of the F test of the excluded instruments is equal to 124.06, which is much larger than the rule of thumb of 10 for strong instruments, and the instruments pass both the Under-identification test and the Weak identification test. The results of the latter two tests are available from the author.

[102] Additionaly, the p-value of the Hansen J test is very high (e.g., p=0.904) for many estimations, which point to the validity of the instruments. From a theoretical point of view, our instruments might be valid since the self-employment status does not directly affect health care utilization [Meer and Rosen (2004)].

[103] Or, more simply, the insurance coverage might have no effect on the LT rehabilitation inpatient days ["LT days" henceforth] since the variability of "LT days" is extremely low: 99.76% of the observations for the "LT days" are equal to zero [Table F1]. The same reasoning might explain why insurance coverage is an exogenous variable (i.e., no reverse causality since "LT days" is mostly composed by zeros).

[104] For example, individuals might forget to turn the steering wheel of the car, implying that the electronic anti-theft device will not work. In this case it would be much easier for a thief to steal a car.

[105] The first behavioral change following the insurance gain (i.e., the ex-ante moral hazard) was tested by Jerant et al. (2013), thus we do not test it in our paper.

[106] Available at `http://www-fars.nhtsa.dot.gov/States/StatesPedestrians.aspx`

gain, then their driving behavior would be affected as well. Thus, we use the pedestrian fatality rate as a proxy for the behavioral carefulness of individuals. We prefer to use the pedestrian fatality rate to other measures (supplied by the NHTSA) such as the car fatality rate for the following reason: the pedestrians cannot cross the federal highways while they can cross local (e.g., city or county) streets, and most of the drivers that use local streets are from that county. On the contrary, most of the drivers that use federal highways are not from that county (e.g., truck drivers), thus if we use the car fatality rate as the dependent variable, we would not identify the effect of the change in the insurance coverage of a county on the drivers' behavior of the same county.

Since the insurance coverage might be endogenous, we instrument the insurance coverage using $Self\ Employment_{it+2}$ and $female\ Obesity_{it}$ and using the FE-IV estimator. The first-stage estimates are reported in Table F12, the reduced form estimates are reported in Table F13, and the second-stage estimates are reported in Table 14. The instruments are strong [Table F12] and we might assume that the instruments are exogenous [see the Hansen J test in Table 14] once we control for unobserved heterogeneity and for many confounders.

Second-Stage IV Estimates

Assuming that our instruments are valid, we find that the insurance coverage is endogenous at the 1% significance level, and an increase of the insurance coverage implies a decrease in the pedestrian fatality rate, thus it seems that individuals drive more carefully after an increase of the insurance coverage.[107] To conclude, an increase of the insurance coverage does not lead to a less careful behavior of the individuals. Thus, the increase of the insurance coverage implies an increase of the larceny and car theft rates through the effect that health insurance has on disposable income (i.e., moral hazard).

8 Conclusion

In this study we find consistent evidence that an increase in the insurance coverage implies an increase in the number of larcenies and car thefts using county level data for the U.S.. These results are found using several types of IV estimators that enable us to disentangle the true effect of the insurance coverage on crime from the effect of endogeneity bias. Additionally, using the (more efficient) GMM estimator we find that an increase of the insurance coverage increases the number of robberies. We relate the previous results with the moral hazard effect that is caused by the insurance gain [e.g., Finkelstein et al. (2012)]. Given our results, the recent implementation of the ACA (Affordable Care Act) -also known as Obamacare- might cause a higher number of (per-capita) larcenies and car thefts.

[107]Also using the Within Group estimator we find a negative effect of the insurance coverage on the dependent variable but the coefficient of coverage is closer to zero in the case of the WG estimator (i.e., -0.258) than using the FE-IV estimator (i.e., -3.221).

References

Altonji, J.G., Elder, T.E. and Taber, C.R. (2005). "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools", *Journal of Political Economy, 113(1)*, pp. 151-184.

American Diabetes Association, (2008). "Diagnosis and Classification of Diabetes Mellitus", *Diabetes Care 31, Supplement 1*: pp. S55-S60.

Anderson, G.F. and Poullier, J.P. (1999). "Health Spending, Access, And Outcomes: Trends In Industrialized Countries", *Health Affairs, 18(3)*: pp. 178-192.

Arrow, K.J. (1968). "The Economics of Moral Hazard: Further Comment", *The American Economic Review, 58(3), Part 1*: pp. 537-539.

Barro, R.J. (1974). "Are Government Bonds Net Wealth?", *Journal of Political Economy, 82(6)*: pp. 1095-1117.

Bayer, P., Hjalmarsson, R. and Pozen, D. (2009). "Building Criminal Capital Behind Bars: Peer Effects in in Juvenile Correction", *Quarterly Journal of Economics, 124 (1)*: pp. 105-147.

Becker, G. (1968). "Crime and punishment: An economic approach", *Journal of Political Economy, 76*: pp. 169-217.

Becker, G. (1993). "Nobel Lecture: The Economic Way of Looking at Behavior", *Journal of Political Economy, 101(3)*: pp. 385-409.

Bellows, J. and Miguel, E.. 2008. War and Local Collective Action in Sierra Leone, Working paper version, 2008. Available at `http://eml.berkeley.edu/~emiguel/pdfs/miguel_sl.pdf`

Blau, D.M. (1987). "A Time-Series Analysis of Self-Employment in the United States", *Journal of Political Economy, 95(3)*: pp. 445-467.

Borjas, G.J. (2000). *"Labor Economics, Second Edition"*, McGraw-Hill Higher Education.

Bundorf, M.K. and Pauly, M.V. (2006). "Is health insurance affordable for the uninsured?" *Journal of Health Economics, 25*: pp. 650-673.

Carpiano, R.M. (2007). "Neighborhood social capital and adult health: An empirical test of a Bourdieu-based model", *Health & Place, 13*: pp. 639-655.

Catlin, A., Cowan, C., Heffler S. and Washington, B.J. (2007). "National Health Spending In 2005: The Slowdown Continues", *Health Affairs, 26(1)*: pp. 142-153.

Catlin, A., Cowan, C., Hartman, M. and Heffler S. (2008). "National Health Spending In 2006: A Year Of Change For Prescription Drugs", *Health Affairs, 27(1)*: pp. 14-29.

Chernew, M., Cutler, D.M. and Keenan, P.S. (2005). "Increasing Health Insurance Costs and the Decline in Insurance Coverage", *Health Services Research, 40(4)*: pp. 1021-1039.

Choe, J. (2008). "Income inequality and crime in the United States", *Economics Letters, 101*: pp. 31-33.

Chiu, W.H. and Madden, P. (1998). "Burglary and income inequality", *Journal of Public Economics, 69*: pp. 123-141.

Cohen, R.A., Makuc, D.M., Bernstein, A.B., Bilheimer, L.T. and Powell-Griner, E. (2009) "Health Insurance Coverage Trends, 19592007: Estimates from the National Health Interview Survey", *National Health Statistics Report, No. 17, 2009.* Available at `http://www.cdc.gov/nchs/data/nhsr/nhsr017.pdf`.

Cullen, J.B. and Levitt, S.D. (1996). "Crime, Urban Flight, and the Consequences for Cities", *NBER, Working Paper 5737.*

Dafny, L.S. (2005). "Are health insurance markets competitive?", *American Economic Review, 100*: pp. 1399-1431.

Dahl, G., and DellaVigna, S. (2009). "Does movie violence increase violent crime?", *The Quarterly Journal of Economics 124(2)*: pp. 677-734.

Dahlberg, M. and Gustavsson, M. (2008). "Inequality and Crime: Separating the Effects of Permanent and Transitory Income", *Oxford Bulletin of Economics and Statistics, 70(2)*: pp. 129-153.

Di Matteo, L. (2005). “The macro determinants of health expenditure in the United States and Canada: assessing the impact of income, age distribution and time”, *Health policy, 71*: pp. 23-42.

Dong, Y. (2013). “How Health Insurance Affects Health Care Demand-A Structural Analysis Of Behavioral Moral Hazard And Adverse Selection”, *Economic Inquiry, 52(2)*: pp. 1324-1344.

Donohue, J.J., and J. Levitt, S.J. (2001). “The Impact of Legalized Abortion on Crime”, *Quarterly Journal of Economics, 116 (2)*: pp. 379420.

Dwyer-Lindgren, L., Freedman, G., Engell, R.E., Fleming, T.D., Lim, S.S., Murray, C.J.L. and Mokdad, A.H. (2013). “Prevalence of physical activity and obesity in US counties, 20012011: a road map for action”, *Population Health Metrics, 11(7)*: pp. 1-11. doi:10.1186/$1478 - 7954 - 11 - 7$.

Ehrlick, I. (1973). “Participation in Illegitimate Activities: A Theoretical and Empirical Investigation”, *Journal of Political Economy, 81(3)*: pp. 521-565.

Ehrlick, I. (1975). “On the Relation between Education and Crime”, in *“Education, Income, and Human Behavior”, NBER*: pp. 313-338.

Eide, E., Rubin, P.H. and Shepherd, J.,M. (2006) “Economics of Crime”, *Foundations and Trends in Microeconomics, now Publishers Inc., Hanover, Massachussets, USA*.

Emanuel, E.J. and Fuchs, V.R. (2004). “Who Really Pays for Health Care? The Myth of Shared Responsibility”, *JAMA, Journal of the American Medical Association, 299(9)*: pp. 1057-1059.

Engstrom, P. and Holmlund, B. (2010). “Tax evasion and self-employment in a high-tax country: evidence from Sweden”, *Applied Economics, 41*: pp. 2419-2430.

Ericson, K. M. and Starc., A. (2012). “Heuristics and Heterogeneity in Health Insurance Exchanges: Evidence from the Massachusetts Connector”, *American Economic Review, 102(3)*: pp. 493-497.

Evans, W.N. and Topoleski, J.H. (2002). “The Social and Economic Impact of Native American Casinos”, *NBER, Working Paper 9198*

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H. and Baicker, K. (2012). “The Oregon health insurance experiment: evidence from the first year”, *The Quarterly Journal of Economics, 127(3)*: pp. 1057-1106.

Fagan, J. and Freeman, R.B. (1999). “Crime and Work”, *Crime and Justice, 25*: pp. 225-290.

Folster, S. (2002). “Do Lower Taxes Stimulate Self-Employment?”, *Small Business Economics, 19*: pp. 135-145.

Jeffrey, R.W. and French, A. (1996). “Socioeconomic Status and Weight Control Practices among 20- to 45-Year-Old Women”, *American Journal of Public Health, 86(7)*: pp. 1005-1010.

Fuchs, V.R. (2000). “The future of health economics”, *Journal of Health Economics, 19*: pp. 141-158.

Gang, I.N. and Zimmermann, K.F. (2000). “Is Child like Parent? Educational Attainment and Ethnic Origin”, *The Journal of Human Resources, 35(3)*: pp. 550-569.

Glaeser, E.L. and Sacerdote, B. (1999). “Why Is There More Crime in Cities?”, *Journal of Political Economy, 107(S6)*: pp. S225-S258.

Goldstein, P.J. (1998). “Drugs, Violence, and Federal Funding: A Research Odyssey”, *Substance Use & Misuse, 33(9)*: pp. 1915-1936.

Grogger, J. (1998). ”Market Wages and Youth Crime”, *Journal of Labor Economics, 16(4)*: pp. 756-791.

Gronau, R. (1977). “Leisure, Home Production, and Work–the Theory of the Allocation of Time Revisited”, *Journal of Political Economy, 85(6)*: pp. 1099-1123.

Glaeser, E.L. and Sacerdote, B. (1999). “Why Is There More Crime in Cities?”, *Journal of Political Economy, 107(S6)*: pp. S225-S258.

Gruber, J. (2001). “Taxes and health insurance”, *NBER Working Paper No. 8657*. Available at `http://www.nber.org/chapters/c10862.pdf`

Gruber, J. (1994). "The Incidence of Mandated Maternity Benefits", *American Economic Review, 84*: pp. 622-641.

Gruber, J. (2000). "Health insurance and the labor market", in: Culyer AJ, Newhouse JP, eds. *Handbook of Health Economics, 1*. New York, NY: Elsevier Science. Available at `http://economics.mit.edu/files/69`

Gruber, J. and Krueger, A.B. (1990). "The Incidence of Mandated Employer-Provided Insurance: Lessons from Workers' Compensation Insurance", *NBER Working Paper No. 3557*

Gould, E.D., Weinberg, B.A. and Mustard, D.B. (2002). "Crime Rates and Local Labor Market Opportunities in the United States: 1979-1997", *The Review of Economics and Statistics, 84(1)*: pp. 45-61.

Hadley, J., Holahan, J., Coughlin, T. and Miller, D., (2008) "Covering The Uninsured In 2008: Current Costs, Sources Of Payment, And Incremental Costs", *Health Affairs, 27(5)*: pp. w399-w415.

Hartman, M., Martin, A., McDonnell, P. and Catlin, A. (2009). "National Health Spending In 2007: Slower Drug Spending Contributes To Lowest Rate", *Health Affairs, 28(1)*: pp. 246-261.

Hartman, M., Martin, A., Nuccio, O., Catlin, A. (2010). "Health Spending Growth At A Historic Low In 2008", *Health Affairs, 29(1)*: pp. 147-155.

Harries, K. (2006). "Property Crimes and Violence in United States: An Analysis of the influence of Population density", *International Journal of Criminal Justice Sciences, 1(2)*: pp. 24-34.

Hayashi, F. (2000). *"Econometrics",* Princeton University Press.

Heineke, J.M. (1978). "Chapter one: Economic models of criminal behavior: an overview", in *Economic models of criminal behavior, North-Holland Publishing Company, 1978.* Available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1908328`

Holahan, J. (2010). "The 2007-2009 Recession And Health Insurance Coverage", *Health Affairs, 30(1)*: pp. 145-152.

Jerant, A., Fiscella, K., Tancredi, D.J. and Franks, P. (2013) "Health Insurance Is Associated With Preventive Care but Not Personal Health Behaviors", *Journal of the American Board of Family Medicine, 26(6)*: pp. 759-767.

Karchmer, C.L., Walsch, M.E., and Greenfield, J. (1979). "Enforcement manual: Approaches for Combatting Arson-for-Profit Schemes (Vol I)", *Washington, DC: U.S. Department of Justice.*

Kasper, J.D., Giovannini, T.A. and Hoffman, C. (2000). "Gaining and Losing Health Insurance: Strengthening the Evidence for Effects on Access to Care and Health Outcomes", *Medical Care Research and Review, 57(3)*: pp. 298-318.

Kemper, P. and Murtaugh, C.M. (1991). "Lifetime Use of Nursing Home Care", *The New England Journal of Medicine, 324*: pp. 595-600.

Klinenberg, E. (2001). "Dying Alone: The Social Production of Urban Isolation", *Etnography, 2*: pp. 501-531.

Kreider, B. and Hill, S.C.(2009). "Partially Identifying Treatment Effects with an Application to Covering the Uninsured", *The Journal of Human Resources, 44(2)*: pp. 409-449.

Lee, D.S. and McCrary, J. (2005). "Crime, punishment, and Myopia", *NBER Working Paper no. 11491.*

Levine, J.A. (2011). "Poverty and Obesity in the U.S.", *Diabetes, 60(11)*: pp. 2667-2668.

Levitt, S.D. (2002). "Using Electoral Cycles in Police Hiring to Estimate the Effects of Police on Crime: Reply", *The American Economic Review, 92(4)*: pp. 1244-1250.

Levitt, S.D. (2004). "Understanding Why Crime Fell in the 1990s: Four Factors That Explain the Decline and Six That Do Not", *The Journal of Economic Perspectives, 18(1)*, pp. 163-190.

Lochner, L. and Moretti, E. (2004). "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review, 94(1)*: pp. 155-189.

Machin, S. and Costas, C. (2004). "Crime and Economic Incentives", *The Journal of Human Resources, 39(4)*: pp. 958-979.

Madrian, B.C. (1994). "Employment-Based Health Insurance and Job Mobility: Is There Evidence of Job-Lock?", *The Quarterly Journal of Economics, 109(1)*: pp. 27-54.

Manski, C.F. (1990). “Nonparametric Bounds on Treatment Effects”, *The American Economic Review, 80(2)*: pp. 319-323.

Martin, A., Lassman, D., Whittle, L. and Catlin, A. (2011). “Recession Contributes To Slowest Annual Rate Of Increase In Health Spending In Five Decades”, *Health Affairs, 30(1)*: pp. 11-22.

Martin, A., Lassman, D., Whittle, L. and Catlin, A. (2012). “Recession Contributes To Slowest Annual Rate Of Increase In Health Spending In Five Decades”, *Health Affairs, 31(1)*: pp. 208-219.

McAuley, W.J., Pecchioni, L.L. and Grant,J.A. (2002). “The Impact of Living in a Rural County With No Nursing Home on Utilization Rates and Admission Mobility Patterns”, *The Journal of Applied Gerontology, 21(1)*: pp. 40-57.

Meer, J. and Rosen, H.S. (2004). “Insurance and the utilization of medical services”, *Social Science & Medicine, 58*: pp 1623-1632.

Imrohoroglu, A., Merlo, A. and Rupert, P. (2004). “What Accounts for the Decline in Crime?”, *International Economic Review, 45(3)*: pp. 707-729.

Miller, R.D. Jr. (2004). “Estimating the Compensating Differential for Employer-Provided Health Insurance”, textitInternational Journal of Health Care Finance and Economics, 4: pp. 27-41.

Miller, S. (2012). “The effect of insurance on emergency room visits: An analysis of the 2006 Massachusetts health reform”, *Journal of Public Economics, 96*: pp. 893-908.

Moore, J. (2005). “Public Health behind Bars: Health Care for Jail Inmates”, *Popular Government, 71(1)*: pp. 16-24. Available at `http://sogpubs.unc.edu/electronicversions/pg/pgfal05/article2.pdf`

Mocan, H.N. and Bali, T.G. (2010). “Asymmetric crime cycles”, *The Review of Economics and Statistics, 92(4)*: pp. 899911.

Murray, M.P. (2006). “Avoiding Invalid Instruments and Coping with Weak Instruments”, *Journal of Economic Perspectives, 20(4)*: pp. 111-132.

Narayan, P.K. and Smith, R. (2004). “Crime rates, male youth unemployment and real income in Australia: evidence from Granger causality tests”, *Applied Economics, 36*: pp. 20792095.

Nunn, N. and Wantchekon, L. (2012). “The Slave Trade and the Origins of Mistrust in Africa”, *American Economic Review, 101(7)*: pp. 3221-3252.

Phelps, C.E. (2003). *“Health Economics”*, Addison Wesley, Pearson Education Ltd.

Okunade, A.A. and Murthy, V.N.R. (2002). “Technology as a ‘major driver’ of health care costs: a cointegration analysis of the Newhouse conjecture”, *Journal of Health Economics, 21*: pp. 147-159.

Pettiway, L.E. (1988). “Urban spatial structure and incidence of arson: Differences between ghetto and nonghetto environments”, *Justice Quarterly, 5(1)*: pp. 113-129.

Raphael, S. and Winter-Ebmer, R. (2001), “Identifying the Effect of Unemployment on Crime”, *Journal of Law and Economics, 44(1)*: pp. 259-283.

Rasanen, P., Hirvenoja, R., Hakko, H. and Isanen, E.V. (1994). “Cognitive functioning ability of arsonists”, *The Journal of Forensic Psychiatry, 5(3)*: pp. 615-620.

Reece, W.S. (2010). “Casinos, hotels and crimes”, *Contemporary Economic Policy, 28(2)*: pp. 145-161.

Robinson, J.C. (2004). “Consolidation And The Transformation Of Competition In Health Insurance”, *Health Affairs, 23(6)*: pp. 11-24.

Rosenthal, S.S. and Ross, A. (2009). “Violent crime, entrepreneurship, and cities”, *Journal of Urban Economics, 67*: pp. 135-149.

Sampson, R.J., Raudenbush, S.W. and Earls, F. (1997). “Neighborhoods and violent crime: a multilevel study of collective efficacy”, *Science, 277(5328)*: pp. 918-924.

Schaenman, P., Davies, E., Jordan, R. and Chakraborty, R. (2013). *“Opportunities for Cost Savings in Corrections Without Sacrificing Service Quality: Inmate Health Care”*, The Urban Institute, Washington, DC. Available at `http://www.urban.org/UploadedPDF/412754-Inmate-Health-Care.pdf?RSSFeed=UI_Governing.xml`

Scorzafave, L.G. and Soares M.K. (2009). “Income inequality and pecuniary crimes”, *Economics Letters, 104*: pp. 40-42.

Shigeoka, H. (2014). “The Effect of Patient Cost Sharing on Utilization, Health, and Risk Protection”, *American Economic Review, 104(7)*: 2152-2184.

Stuart, H. (2003). “Violence and mental illness: an overview”, *World Psychiatry, 2(2)*: pp. 121-124.

Vitaro, F., Brendgen, M., Ladouceur, R. and Tremblay, R.E. (2001). “Gambling, Delinquency, and Drug Use During Adolescence: Mutual Influences and Common Risk Factors”, *Journal of Gambling Studies, 17(3)*: pp. 171-190.

Ward, L. and Franks, P. (2007).“Changes in Health Care Expenditure Associated with Gaining or Losing Health Insurance”, *Annals of Internal Medicine, 146(11)*: pp. 768-774.

Zweifel, P. and Manning, W.G. (2000). “Chapter 8 Moral hazard and consumer incentives in health care”, *Handbook of Health Economics, 1(A)*: pp. 409-459.

Western, B., Kleykamp M. and Rosenfeld J. (2006). “Did Falling Wages and Employment Increase U.S. Imprisonment?”, *Social Forces, 84(4)*: pp. 2291-2311.

Windmeijer, F. (2005). “A finite sample correction for the variance of linear efficient two-step GMM estimators”, *Journal of Econometrics, 126*: pp. 25-51.

Wooldridge, J.M. (2010). *“Econometric Analysis of Cross Section and Panel Data”*, Cambridge, MA: MIT Press.

Zhang, J. (1997). “The Effect of Welfare Programs on Criminal Behavior: A Theoretical and Empirical Analysis”, *Economic Inquiry, 35(1)*: pp. 120-137.

Table 1: Descriptive Statistics: Panel structure

Year	(1) Observations	(2) Percentage	(3) Cumulative Percentage
2005	2843	16.94	16.94
2006	2810	16.74	33.68
2007	2797	16.66	50.34
2008	2795	16.65	66.99
2009	2758	16.43	83.42
2010	2781	16.57	100
All years	16784	100	100

Table 2: Descriptive Statistics: Insurance Coverage and Crimes (Crimes per 1000 individuals)

	(1) Mean	(2) Standard Deviation
Insurance coverage rate	0.817	0.058
Crimes per 1000 persons:		
Larceny	13.882	9.661
Car theft	1.340	1.378
Burglary	5.288	3.894
Arson	0.144	0.224
Robbery	0.351	0.572
Aggrav. assault	1.826	1.762
Murder	0.030	0.057
Rape	0.232	0.237
Total Observations ($N \times T$)	16784	

Notes: Variables are normalized per 1000 individuals. Sources: NACJD and SAHIE.

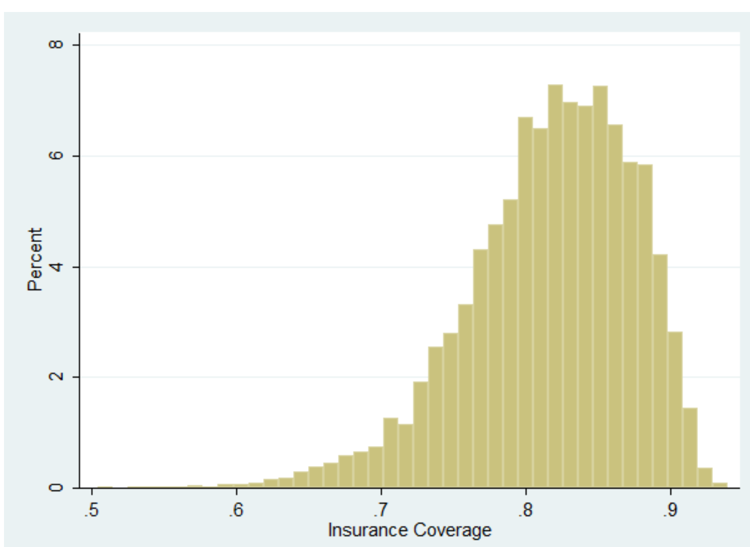


Figure 8-3: The distribution of insurance coverage

Table 3: Descriptive Statistics: Controls

	(1) Mean	(2) Standard Deviation
White-Non-Hispanic (Fraction)	0.797	0.191
White-Hispanic (Fraction)	0.073	0.126
Black-Non-Hispanic (Fraction)	0.083	0.139
Black-Hispanic (Fraction)	0.002	0.003
Asian-Non-Hispanic (Fraction)	0.010	0.020
Asian-Hispanic (Fraction)	0.000	0.001
Other-Non-Hispanic (Fraction)	0.030	0.071
Other-Hispanic (Fraction)	0.005	0.006
Unemployment (%)	6.502	3.004
Poverty (%)	15.718	6.225
Male (Fraction)	0.499	0.021
Population 15-19 (Fraction)	0.072	0.011
Population 20-24 (Fraction)	0.061	0.025
Population >64 (Fraction)	0.156	0.041
Social Security recipients (Fraction)	0.208	0.050
Population density	152.676	567.815
# Police forces ($t-1$)	113.830	429.114
Personal Income (in 2005 $)	29636.064	6847.411
Small Firms (Fraction)	0.915	0.036
Gambling arrests (per 1000 persons)	0.016	0.169
Drug arrests (per 1000 persons)	4.315	7.513
Total Observations ($N \times T$)	16784	

Notes: Sources: U.S. Census Bureau, BEA, FBI, NACJD.

Table 4: (Pooled) Property Crimes VS (Pooled) Violent Crimes (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes§ per 1000 individuals				Violent Crimes§ per 1000 individuals			
Insurance Coverage	6.648*	6.429*	8.300***	10.058***	-3.728***	-3.767***	-0.586	-0.877
	(3.680)	(3.694)	(3.146)	(3.306)	(0.572)	(0.574)	(0.586)	(0.659)
Observations	18492	18492	18492	16784	18492	18492	18492	16784
Adj. R-squared	0.0007	0.0017	0.8752	0.8901	0.0088	0.0092	0.833	0.827
YEAR FE	NO	YES	YES	YES	NO	YES	YES	YES
COUNTY FE	NO	NO	YES	YES	NO	NO	YES	YES
Controls†	NONE	NONE	NONE	ALL	NONE	NONE	NONE	ALL

Notes: $***p < 0.01$, $**p < 0.05$ $*p < 0.10$. Clustered Standard errors at the county level in parenthesis. The number of reported crimes by county are expressed per 1000 individuals and the insurance coverage is a fraction ranging from 0.505 to 0.967 at the county level. The unit of analysis is the county.

§: Property Crimes include: larcenies, burglaries and motor-vehicle thefts, arsons. Violent crimes include: murders, rapes, aggravated assaults and robberies.

†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year ($t-1$), 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table 5: WG Estimator: Breakdown by type of Crime (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
Ins. Coverage	9.552***	1.120**	-0.525	-0.090	0.014	-0.051	-0.548	-0.292***
	(2.125)	(0.439)	(1.377)	(0.167)	(0.029)	(0.119)	(0.587)	(0.108)
Observations	16784	16784	16784	16784	16784	16784	16784	16784
Adj. R-Squared‡	0.887	0.849	0.824	0.359	0.200	0.529	0.781	0.914
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis. The number of reported crimes by county are expressed per 1000 individuals and the insurance coverage is a fraction ranging from 0.505 to 0.967 at the county level. The unit of analysis is the county.

†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

‡: The Adjusted R-squared includes the effect of the County Fixed Effects, and is found using the command areg in Stata 12.

Table 6: Reduced-Form Regressions (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
$Self\ Employment_{it+2}$	4.583*	0.952*	2.257	0.113	-0.059	-0.044	0.118	-0.149
	(2.460)	(0.491)	(1.920)	(0.217)	(0.048)	(0.124)	(0.671)	(0.106)
Observations	16772	16772	16772	16772	16772	16772	16772	16772
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis.

†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table 7: FE-IV Regressions: Second stage estimates (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
Ins. Coverage	55.448*	11.517**	27.310	1.373	-0.717	-0.539	1.435	-1.806
	(29.385)	(5.764)	(22.424)	(2.412)	(0.554)	(1.378)	(7.385)	(1.273)
Hausman test (Coverage)§:								
Test statistic: $\chi_1^2 =$	2.933	3.863	1.824	0.402	1.852	0.132	0.074	1.576
p-value	***0.087***	***0.049***	0.177	0.526	0.174	0.716	0.786	0.209
Observations	16772	16772	16772	16772	16772	16772	16772	16772
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: $***p < 0.01$, $**p < 0.05$ $*p < 0.10$. Clustered Standard errors at the county level in parenthesis. The number of reported crimes by county are expressed per 1000 individuals and the insurance coverage is a fraction ranging from 0.505 to 0.967 at the county level. The unit of analysis is the county.

§: The instrument is: 1) $Self\ Employment_{it+2}$.

†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (<500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table 8: FE-IV Regressions: Second stage results (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
Ins. Coverage	56.786**	8.857**	18.863	0.065	-0.464	-1.387	4.553	-0.929
	(22.539)	(4.189)	(15.592)	(1.727)	(0.372)	(1.049)	(5.603)	(0.955)
Hausman test (Coverage)								
Test statistic: $\chi_1^2 =$	5.403	3.016	1.007	1.812	0.644	3.476	1.101	0.418
p-value	***0.020***	***0.082***	0.316	0.178	0.422	***0.062***	0.294	0.518
Hansen J statistic§: $\chi_1^2 =$	0.007	0.967	0.806	2.464	1.142	1.575	0.568	1.477
p-value	0.936	0.325	0.369	0.117	0.285	0.209	0.451	0.224
Observations	16772	16772	16772	16772	16772	16772	16772	16772
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: $***p < 0.01$, $**p < 0.05$ $*p < 0.10$. Clustered Standard errors at the county level in parenthesis. The number of reported crimes by county are expressed per 1000 individuals and the insurance coverage is a fraction ranging from 0.505 to 0.967 at the county level. The unit of analysis is the county.

§: The instruments are: 1) $Self\ Employment_{it+2}$; 2) $Diabetes\ Incidence_{it-1}$.

†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (<500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table 9: FE-IV Regressions: Second stage results (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
Ins. Coverage	42.270***	7.216***	0.240	0.089	-0.074	-0.196	1.265	-0.054
	(14.149)	(2.576)	(8.342)	(1.075)	(0.187)	(0.577)	(3.591)	(0.648)
Hausman test (Coverage)								
Test statistic: $\chi^2_1 =$	6.244	6.661	0.582	0.075	0.150	0.036	0.684	0.038
p-value	***0.012***	***0.010***	0.445	0.783	0.699	0.849	0.408	0.845
Hansen J statistic§: $\chi^2_1 =$	1.369	3.727	3.877	6.294	2.671	6.240	2.591	2.757
p-value	0.849	0.444	0.423	0.178	0.614	0.182	0.628	0.599
Observations	16772	16772	16772	16772	16772	16772	16772	16772
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$ $^{*}p < 0.10$. Clustered Standard errors at the county level in parenthesis. The number of reported crimes by county are expressed per 1000 individuals and the insurance coverage is a fraction ranging from 0.505 to 0.967 at the county level. The unit of analysis is the county.
§: The instruments are: 1) $Self\ Employment_{it+2}$; 2) $Diabetes\ Incidence_{it-1}$; 3) $Female\ Obesity_{it}$; 4) $Male\ Obesity_{it}$; 5) $Female\ Obesity_{it-3}$.
†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table 10: FD-IV Regressions: Second stage results (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
Ins. Coverage	34.485***	7.377**	-15.188	1.532	-0.244	1.023	0.987	0.313
	(12.887)	(3.612)	(10.354)	(1.241)	(0.218)	(0.706)	(4.248)	(0.648)
Hausman test (Coverage)								
Test statistic	4.576	3.846	1.638	3.581	1.310	1.677	0.086	0.849
χ^2_1 p-value	***0.032***	***0.050***	0.201	0.058	0.252	0.195	0.770	0.357
Hansen J statistic§: $\chi^2_3 =$	2.001	4.235	1.818	4.233	0.670	1.276	4.302	6.614
p-value	0.572	0.237	0.611	0.237	0.880	0.734	0.231	0.085
Observations	13418	13418	13418	13418	13418	13418	13418	13418
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$ $^{*}p < 0.10$. Clustered Standard errors at the county level in parenthesis. The number of reported crimes by county are expressed per 1000 individuals and the insurance coverage is a fraction ranging from 0.505 to 0.967 at the county level. The unit of analysis is the county.
§: The instruments are: 1) $\Delta Hospital\ Days\ Retarded_{it}$, 2) $\Delta Home\ health\ agencies_{it}$, 3) $\Delta Female\ obesity_{it}$, 4) $Self\ employment_{it}$.
†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table 11: Difference GMM Estimations (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
	One-step GMM estimator‡							
Coverage	27.341**	5.119***	-6.967	-0.778	-0.320	-0.204	4.980	2.780**
	(10.885)	(1.791)	(5.993)	(0.919)	(0.220)	(1.202)	(5.181)	(1.078)
	Two-step GMM estimator‡‡							
Coverage	19.123**	3.810**	-3.749	-0.277	-0.128	-1.097	3.450	1.793*
	(9.602)	(1.563)	(4.931)	(0.570)	(0.208)	(1.178)	(4.720)	(0.942)
Hansen J test								
Statistic§:	$\chi^2_5 = 5.41$	$\chi^2_5 = 5.17$	$\chi^2_5 = 10.79$	$\chi^2_5 = 4.12$	$\chi^2_5 = 10.08$	$\chi^2_7 = 9.21$	$\chi^2_7 = 8.08$	$\chi^2_7 = 10.46$
P-value	0.368	0.395	0.056	0.532	0.073	0.238	0.326	0.164
Observations	13433	13433	13433	13433	13433	13433	13433	13433
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$.
‡: Robust standard errors (at the county level), consistent in the presence of any pattern of heteroskedasticity and autocorrelation within panels, in parenthesis.
‡‡: Robust standard errors (at the county level), consistent in the presence of any pattern of heteroskedasticity and autocorrelation within panels, with Windmeijer (2005)'s finite-sample correction for the two-step covariance matrix, in parenthesis.
The number of reported crimes by county are expressed per 1000 individuals and the insurance coverage is a fraction ranging from 0.505 to 0.967 at the county level. The unit of analysis is the county.
§: The instruments for Columns (1)-(5) are: 1) $\Delta Coverage_{it-3}$, 2) $Coverage_{it+2}$, 3) $Coverage_{it+3}$, 4) $Self\ Employment_{it-2}$, 5) $Self\ Employment_{it-3}$, 6) $\Delta Retardation_{it-3}$. The instruments for Columns (6)-(8) are: 1) $\Delta Self\ Employment_{it-3}$, $\Delta Self\ Employment_{it-4}$, 3) $\Delta Retardation_{it-3}$, 4) $\Delta Retardation_{it-4}$, 5) $\Delta Female\ obesity_{it-3}$, 6) $\Delta Female\ obesity_{it-4}$, 7) $Coverage_{it+3}$, 8) $Coverage_{it+4}$.
†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table 12: The effect of insurance coverage on (per-capita) hospital and nursing home utilization, 2^{nd} stage regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Hospital utilization rate	Total Hosp. Inpatient days (per capita)	STG Inpat. Hosp. days (per capita)	STNG-LT Inpat. Hosp. days (per capita)	ST Community H. Inp. days (per capita)	ST Rehab. Inpat. days (per capita)	LT Rehab. Inpat. days (per capita)	Total Nursing Home Inp.days(p.c.)
Health Ins. Coverage	4.860***	18.117***	13.012***	5.104***	13.134***	0.254***	0.002	4.785***
	(0.555)	(2.441)	(1.956)	(1.192)	(1.905)	(0.101)	(0.006)	(1.392)
Hansen J statistic§	0.408	0.014	0.115	0.248	0.110	2.215	2.027	1.373
χ^2_1 P-value	0.522	0.904	0.735	0.618	0.739	0.136	0.154	0.241
Hausman test statistic	52.735	54.712	45.603	18.111	50.734	4.609	0.202	15.366
χ^2_1 P-value	***0.000***	***0.000***	***0.000***	***0.000***	***0.000***	***0.031***	0.653	***0.000***
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
STATE-TIME DUMMIES	YES	YES	YES	YES	YES	YES	YES	YES
CONTROLS†	YES	YES	YES	YES	YES	YES	YES	YES
Observations	18474	18474	18474	18474	18474	18474	18474	18474

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$. Clustered Standard errors at the county level in parenthesis. State FE, Year Dummies are included in each regression.
†: The controls are: 8 race dummies (where the baseline is White and Non Hispanic), sex, 4 dummies for the age distribution in each county, population density.
§: The two instruments used for the insurance coverage are $Self\ Employment_{i,t-1}$ and $Self\ Employment^{Neigh.}_{i,t}$ (the self employment in the neighboring county).

Table 13: The effect of insurance coverage on (per-capita) Outpatient Hospital visits, 2^{nd} stage regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	STG Nursing home Inp. days (per capita)	Total Hosp. Nursing home Admissions (per capita)	STG Hosp. Nursing home Admissions (per capita)	STNGLT Hosp. Nursing home Admissions (per capita)	LT Total Outpatient Visits (per capita)	LT Clinic Outpatient Visits (per capita)
Health Ins. Coverage	4.549***	0.040***	0.038***	0.002*	0.604*	2.098***
	(1.357)	(0.006)	(0.006)	(0.001)	(0.331)	(0.572)
Hansen J statistic§	2.419	0.120	0.511	2.767	1.452	1.372
χ^2_1 P-value	0.119	0.729	0.474	0.096	0.228	0.241
Hausman test statistic	14.759	41.311	39.386	3.504	3.888	15.696
χ^2_1 P-value	***0.000***	***0.000***	***0.000***	0.061	***0.048***	***0.000***
YEAR FE	YES	YES	YES	YES	YES	YES
STATE-TIME TRENDS	YES	YES	YES	YES	YES	YES
CONTROLS†	YES	YES	YES	YES	YES	YES
Observations	18474	18474	18474	18474	18474	18474

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Clustered Standard errors at the county level in parenthesis. State FE, Year Dummies are included in each regression.
†: The controls are 7 race dummies (where the baseline is White and Non Hispanic), sex, 4 dummies for the age distribution in each county, population density.
§: The two instruments used for the insurance coverage are $Self\ Employment_{i,t-1}$ (i.e., the lagged self employment by county) and $Self\ Employment^{Neigh.}_{i,t}$ (the self employment in the neighboring county).

Table 14: Second-Stage Regression

	(1)
	Dep. Var.: Pedestrian Fatality Rate
Insurance Coverage	-3.221***
	(1.223)
Hansen J Statistic§	0.050
P-value	0.823
Hausman Test Statistic	7.425
P-value	0.006
Observations	16754
YEAR FE	YES
COUNTY FE	YES
CONTROLS†	YES

Notes: *** $p < 0.01$, ** $p < 0.05$ * $p < 0.10$. Clustered Standard errors at the county level in parenthesis.
§: The instruments are: $Self\ Employment_{it+2}$, $Female\ Obesity_{it}$.
†: The list of controls at the county level include the full set of controls: see the notes of Table F5.

A Employer-based Health Insurance and Job-Lock

Another characteristic of the U.S. health insurance markets impacts on workers' wages, and it might indirectly impact on the property crime rates: Madrian (1994) shows that *the gain of employer-based health insurance decreases workers' job mobility* (i.e., job lock), which in turn might decrease the quality of the worker-employer match. Madrian (1994) finds that workers with employer-based health insurance have 25% lower voluntary turnover rate than workers without employer-based health insurance. Thus, an individual that was uninsured and that becomes insured through employer-based insurance might lose better job market opportunities, that may incentivize her (or younger members of her family) to commit property crime.[108]

B Lagrangian maximization

BA Interior solution (Case $N > 0$)

In the case of an interior solution (i.e., $L > 0, C > 0, N > 0, X_M > 0$), the Lagrangian function is given by

$$\mathcal{L} = U((X_M + f(C)), L) + \lambda(WN + V - Benefits(p) - pNetCost - X_M) + \mu(T - L - C - N) \tag{28}$$

and we maximize the Lagrangian function with respect to L, X_M, C and N. The first order conditions are

$$\partial U()/\partial L = \mu \tag{29}$$

$$\partial U()/\partial X_M = \lambda \tag{30}$$

$$\frac{\partial U()}{\partial f(C)} \frac{\partial f(C)}{\partial C} = \mu \tag{31}$$

$$\lambda W = \mu \tag{32}$$

From equations (29)-(30), we obtain $\frac{\partial U()/\partial L}{\partial U()/\partial X_M} = \frac{\mu}{\lambda}$, while from equation (32) we obtain $W = \frac{\mu}{\lambda}$. Since $X_C = f(C)$,[109] and since the individual is indifferent between receiving utility from "crime goods" (X_C) or from "market goods" (X_M)[110] -i.e., $\frac{\partial U()}{\partial X_C} = \frac{\partial U()}{\partial X_M} = \frac{\partial U()}{\partial X}$- we can rewrite (31) as

$$\frac{\partial U()}{\partial X_M} f' = \mu. \tag{33}$$

Using (30) and (33), we finally obtain $f' = \frac{\mu}{\lambda}$. Thus, we arrive to

$$\frac{\partial U/\partial L}{\partial U/\partial X} = f' = W = \frac{\mu}{\lambda} \tag{34}$$

where $\frac{\mu}{\lambda}$ is the shadow price of time (in monetary terms), where μ and λ are the marginal utilities of time and income, respectively.

BB Case $N \geq 0$

In this case the Lagrangian function is given by

$$\mathcal{L} = U((X_M + f(C)), L) + \lambda(WN + V - Benefits(p) - pNetCost - X_M) + \mu(T - L - C - N) + \Psi N \tag{35}$$

where we take into account that $N \geq 0$ [see the CS (Complementary Slackness) Condition in equation (41)]. We maximize the Lagrangian function with respect to L, X_M, C and N, thus the first order conditions are

$$\partial U()/\partial L = \mu \tag{36}$$

$$\partial U()/\partial X_M = \lambda \tag{37}$$

[108] Still, also for this individual we have the moral hazard effect of the insurance gain.
[109] Thus $\frac{\partial U()}{\partial X_C} = \frac{\partial U()}{\partial f(C)}$.
[110] [...] since C and N are perfect substitutes. See Footnote 21.

$$\frac{\partial U()}{\partial f(C)}\frac{\partial f(C)}{\partial C} = \mu \tag{38}$$

$$\lambda W = \mu - \Psi \tag{39}$$

$$N \geq 0 \tag{40}$$

$$CS\ Condition: \quad \Psi N = 0\ and\ \Psi \geq 0 \tag{41}$$

The first order conditions with respect to L, X_M and C [equations (36), (37), (38)] are the same as in Section BA, thus we obtain $\frac{\partial U/\partial L}{\partial U/\partial X} = f' = \frac{\mu}{\lambda}$. From (39) we obtain

$$W = \frac{\mu}{\lambda} - \frac{\Psi}{\lambda}, \tag{42}$$

Since $N \in \mathbb{R}_+$, we have two cases:

- In the case that $N > 0$, by the CS condition we know that $\Psi = 0$, therefore (34) still holds.
- On the contrary, if $N = 0$, by the CS condition we obtain that $\Psi \geq 0$ [see (41)], so we have two cases: if $\Psi = 0$, then (34) still holds but if $\Psi > 0$, then (34) does not hold and we obtain

$$\frac{\partial U/\partial L}{\partial U/\partial X} = f' = \frac{\mu}{\lambda} > W, \tag{43}$$

  thus the wage is lower than the productivity of crime.
  To conclude, in the case that $N = 0$, the necessary conditions for the optimum are more general than in the case that $N > 0$:

$$\frac{\partial U/\partial L}{\partial U/\partial X} = f' = \frac{\mu}{\lambda} \tag{44}$$

C Analysis of the model

To analyze the model, we simplify the non-labor income (V) minus the health care expenses as $V^{\dagger} = V - Benefits(p) - p \times NetCost$, and we write the consumer's marginal rate of substitution (MRS) between goods and leisure $\left(\frac{\partial U/\partial L}{\partial U/\partial X}\right)$ as

$$MRS(X, L) = \frac{U_L\left[X_M + X_C, L\right]}{U_X\left[X_M + X_C, L\right]} = \frac{U_L\left[WN + V^{\dagger} + f(C), T - C - N\right]}{U_X\left[WN + V^{\dagger} + f(C), T - C - N\right]} = ...$$

$$= m[N, V^{\dagger} + f(C), T - C] \tag{45}$$

where $U_L\left[\cdot\right]$ and $U_X\left[\cdot\right]$ are the derivatives of the utility function with respect to L and X, respectively. Reformulating the individual's MRS as in eq. (45) stresses an important characteristic of the model: the individual's choice of criminal hours (C) influences her choice of work hours (N) by changing her effective non-labor income, $V^{\dagger} + f(C)$, and her effective time available, $T - C$.

The individual's reservation wage is defined as $W^* = m(0, V^{\dagger}, T)$: the reservation wage is her MRS evaluated at the point where all of her time is allocated to leisure. *The necessary condition for working* is $W > W^*$ where W is the market wage. Similarly, the *necessary condition for committing crime* is $f'(0) > W^*$ (i.e., the returns to the first hour of crime must be larger than the reservation wage).[111]

An individual that both works in the market and spends time in crime chooses optimal criminal hours to equate the marginal returns of the two activities: $f'(C) = W$ [see eq. (6)] The individual's problem has a recursive structure: first, the individual decides how much time to spend in crime and then she decides how much time to spend working in the market.

Using $f'(C) = W$, we can obtain a *participation rule for criminals who work*. By the concavity of $f(C)$ and using the assumption of diminishing marginal utility, a necessary condition for the individual to commit crime and participate in the labor market is $f'(0) > W$ (i.e., the returns to the first hour of crime must exceed the wage). If this condition is satisfied,[112] then the individual chooses time in crime (C) until $f'(C) = W$, and then she will choose N (i.e., labor working time) to satisfy $W = m[N, V^{\dagger} + f(C), T - C]$.[113]

Figure C-4 illustrates the individual's decisions on working hours (N) and on hours spent in crime (C) for two

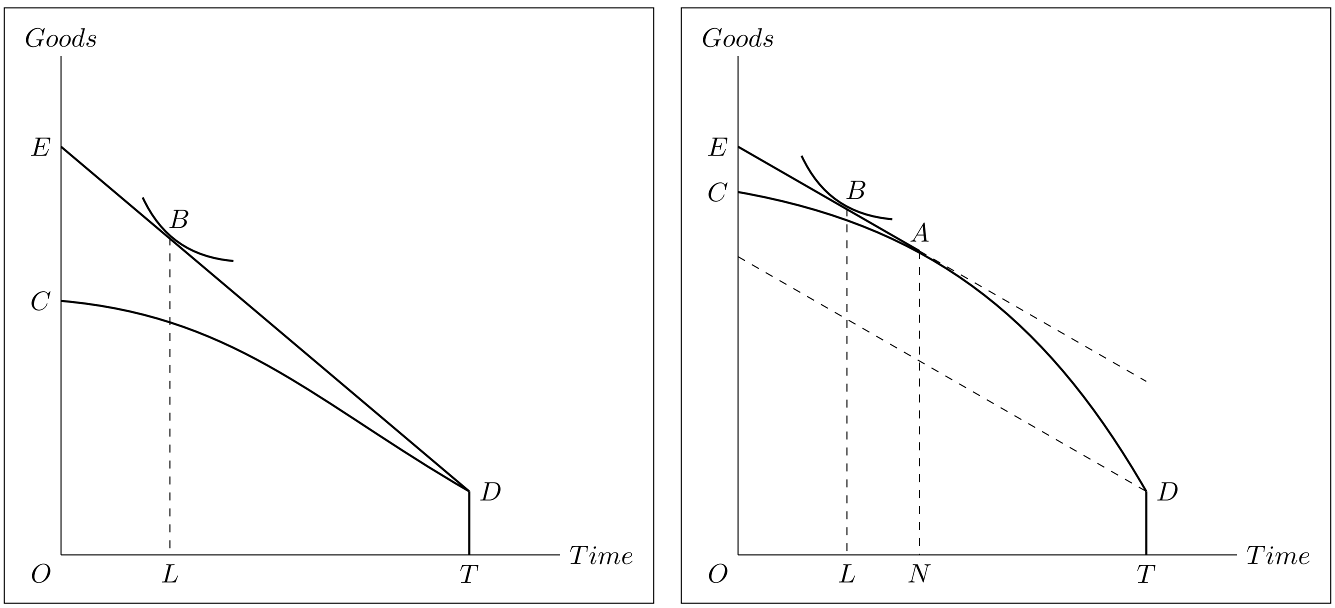
[111] It can happen that the two necessary conditions ($W > W^*$ and $f'(0) > W^*$) are simultaneously not satisfied; for example, an individual with very high non-labor income (e.g., a rentier) tends to have a very high reservation wage (W^*), thus she might both not work and not be involved in crime.

[112] And if $W > W^*$ (i.e., the individual is working in the labor market).

[113] In the case that $f'(0) \leq W$, the individual commits no crime ($C = 0$). Assuming that $W > W^*$ for this individual, then the optimal labor supply (N) is chosen in the usual way, by equating the wage to the MRS, so that $W = m[N, V^{\dagger}, T]$.

individuals who work in the market. Individual A and B both face the same wage, given by (minus) the slope of the budget line, and both have exogenous income $V^\dagger$ [point D in Figure C-4]. Returns to crime, given by $f(C)$, differ between the two individuals: individual A has a very low return to crime while individual B has a high return to crime. Both individuals have the same *preference for leisure versus goods* which is captured by the curved line above the budget set line.[114] For individual A the return to the first hour of crime ($f'(0)$) given by the slope of $f(C)$ at T hours of leisure are lower than the wage, thus she sets $C = 0$ (i.e., zero time units in crime). She chooses market hours N at the point where the wage is equal to the MRS (i.e., $W = m[N, V^\dagger, T]$). Finally, individual A enjoys OL units of leisure time, and spends LT units of time in labor market work.

On the contrary, for individual B the return to the first hour of crime is higher than the wage ($f'(0) > W$). As a result, consumer B spends time in crime, and she chooses C such that $f'(C) = W$ [see point N in Figure C-4]. For this individual the effective non-labor income is $V^\dagger + f(C)$ and the effective time available is $T - C$,[115] then she chooses her optimal labor supply to satisfy $W = m[N, V^\dagger + f(C), T - C]$. To conclude, individual B enjoys OL units of leisure time, spends LN time units in labor market work, and spends NT time units in crime.



(a) INDIVIDUAL A

(b) INDIVIDUAL B

Figure C-4

There is also a third case -the case of Individual C- such that $f'(0) > W$, thus the individual is involved in crime, and the individual has a high preference for leisure, so that the MRS curve is tangent to the crime returns curve ($f(C)$), and the individual offers zero units of labor market work. In this case the individual splits her time into crime and leisure as shown in Figure C-5. The individual enjoys OL' units of leisure and spends $L'T$ units of time in crime.

[114] Also, each individual in this model can have a different preference for leisure: some individual might have low preference for leisure (in this case the interval OL is very small) or they might have high preference for leisure (in this case the interval OL is very large).

[115] I.e., the interval NT is the time in crime.

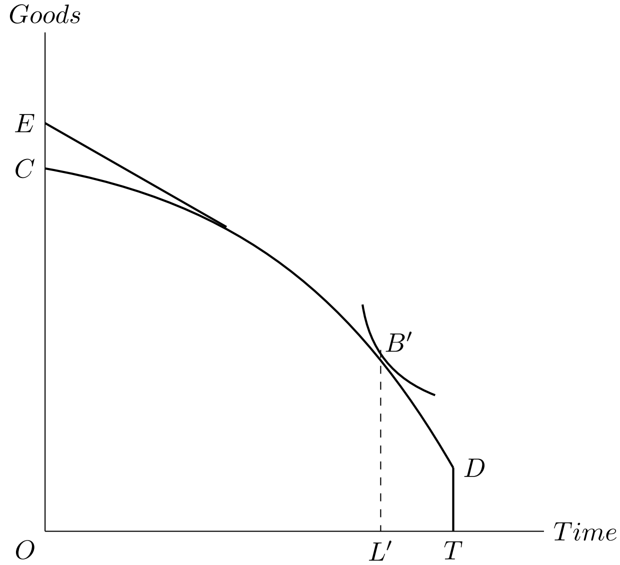


Figure C-5: INDIVIDUAL C

D The Altonji et al. (2005)'s method for omitted variable bias

The parameter of interest is the coefficient (α) of insurance coverage (x), and let q be an (unobservable) index of socio-economic characteristics ["socio-economic index" henceforth] that is correlated with x and with y. We would like to estimate:

$$y = \alpha x + \gamma q + \epsilon \tag{46}$$

Since we cannot control for q, the estimate of α suffers from omitted variable bias:

$$plim \;\; \widehat{\alpha}_{OLS,NC} = \alpha_0 + \gamma \frac{Cov(x,q)}{Var(x)} \tag{47}$$

where NC denotes "No Controls". Suppose that there is a set of observable controls (w)[116] that are directly related to q, and that are related to x only through q:[117]

$$q = w\beta + \tilde{q} \tag{48}$$

where $\tilde{q}$ is an unobservable part of the socio-economic index. Now, our true model is eq. (46), but we observe only one part of q, thus we estimate:

$$y = \alpha x + w\beta + v, \tag{49}$$

then the OLS estimate of (α) will have the following bias:[118]

$$plim \;\; \widehat{\alpha}_{OLS,C} = \alpha_0 + \gamma \frac{Cov(x,\tilde{q})}{Var(x)} \tag{50}$$

where C denotes "Controls". Using eq. (47),(48),(50), The difference between the estimates of α with and without controls is given by:[119]

$$\widehat{\alpha}_{OLS,NC} - \widehat{\alpha}_{OLS,C} \approx \gamma \frac{Cov(x,w\beta)}{Var(x)} \tag{51}$$

From eq. (51) we know that the stronger is the effect of the "socio-economic index" on crime (i.e., γ), or the stronger is the covariance between the insurance coverage and the controls, the larger is the reduction of the bias after we include

[116] E.g., the set of controls include the poverty rate, the personal income and the population density.

[117] This assumption does not imply that $Cov(x,w) = 0$ since highly likely we have that simultaneously $Cov(w,q) \neq 0$ and $Cov(x,w) \neq 0$.

[118] The bias is found inserting eq. (48) inside eq. (46), obtaining $y = \alpha x + \gamma\,[w\beta + \tilde{q}] + \epsilon$, and from the OLS estimation of the latter equation we find the bias expressed in eq. (50).

[119] We use the approximately equal sign ($\approx$) rather than the equality sign since we go from the large sample theory used in eq. (47)-(50) to the small sample theory of the following discussion.

the controls.

Our final goal is to understand how strong the covariance between the unobserved part of the "socio-economic index" and the insurance coverage (i.e., $Cov(x, \tilde{q})$) needs to be in order that the estimate of α using eq. (50) is completely driven by omitted variable bias. Setting $\alpha_0 = 0$ in eq. (50) and dividing eq. (50) by eq. (51), we obtain:

$$\frac{\widehat{\alpha}_{OLS,C}}{\widehat{\alpha}_{OLS,NC} - \widehat{\alpha}_{OLS,C}} \approx \frac{Cov(x, \tilde{q})}{Cov(x, w\beta)} \tag{52}$$

The ratio on the left hand side is obtained with the corresponding regressions with and without controls. The ratio on the right hand side is a (theoretical) measure of how strong the covariance between the unobserved part of the "socio-economic index" and the insurance coverage (i.e., $Cov(x, \tilde{q})$) must be, relative to the covariance between the observed part of the "socio-economic index" and the insurance coverage, in order that the estimation of α is completely driven by omitted variable bias. Thus, a large ratio means that it is implausible that the estimate of α is completely driven by omitted variable bias. Nunn et al. (2012) [pp. 3238] seems to suggest that any ratio above 1 is acceptable (i.e. no omitted variable bias). For example, if the estimated ratio is equal to 2, to attribute the entire OLS estimate of α to omitted variable bias, $Cov(x, \tilde{q})$ would have to be two times greater that $Cov(x, w\beta)$.

E The Health Care Utilization Measures

We use a large set of hospital and nursing home utilization measures as our dependent variables to test for the presence of (ex-post) moral hazard. We use: 1) a general measure of hospital (inpatient) utilization rate, 2) various measures of hospital inpatient days; 3) various measures of nursing home admissions and nursing home inpatient days, 4) two measures of hospital outpatient visits.

A General Measure of Hospital (Inpatient) Utilization Rate

The AHRF dataset contains the distribution of hospitals by utilization rate (for each county).[120] Four random variables are contained in the AHRF dataset (by county): 1) the number of hospitals with 0-39% of utilization rate ["# rate 20%" (see below)], 2) the number of hospitals with 40-59% of utilization rate ["# rate 50%"], 3) the number of hospitals with 60-80% of utilization rate ["# rate 70%"], 4) the number of hospitals with 80%-100% of utilization rate ["# rate 90%"].[121] In order to obtain a single measure for the hospital utilization rate in each county, we created the following random variable:

$$\text{utilization}_{it} = \frac{0.20 * \text{“}\#rate\ 20\%'' + 0.50 * \text{“}\#rate\ 50\%'' + 0.70 * \text{“}\#rate\ 70\%'' + 0.90 * \text{“}\#rate\ 90\%''}{(\text{Number of hospitals})_{it}}$$

where i indicates the county and t indicates the year. If a county does not contain any hospital, we set $\text{utilization}_{it} = 0$.[122] Thus, the random variable Utilization_{it} can range from 0 to 0.90.

Measures of Hospital Inpatient Days

We use the following measures of *per-capita* hospital inpatient days: a) the number of total hospital inpatient days, b) the number of Short Term General (STG) hospital inpatient days,[123] c) the number of Short Term Non General-Long Term (STNG-LT) hospital inpatient days, d) the number of Short Term (ST) Community hospital inpatient days, e) the number of Short Term (ST) rehabilitation hospital inpatient days; f) the number of Long Term (LT) rehabilitation hospital inpatient days.

Measures of Nursing Home Admissions and Nursing Home Inpatient Days

[120] The utilization rate of each hospital at time t is calculated as

$$(\text{Utilization Rate})_{ht} = \frac{(\text{Inpatient days})_{ht}}{365 * (\text{Beds})_{ht}},$$

where $(\text{Inpatient days})_{ht}$ is the number of inpatient days at the hospital h at time t and $(\text{Beds})_{ht}$ is the number of beds of hospital h at time t.

[121] For example, if there are two hospitals in a county and if both hospitals have a utilization rate of 40-59%, then the second random variable [see 2) above] is equal to 2, while the other three random variables are equal to 0.

[122] $\text{Utilization}_{it} = 0$ for 620 counties out of 3141 counties in each year during 2005-2010, since these counties are not served by any hospital.

[123] Each hospital is categorized based on whether the hospital is treating short term (ST) diseases or long term (LT) diseases, and whether the hospital is treating general (G) diseases or non-general (NG) diseases.

We use three measures of *per-capita* nursing home admissions: a) total nursing home admissions, b) STG nursing home admissions, c) STNG-LT nursing home admissions. We use two measures of *per-capita* nursing home inpatient days: a) total nursing home inpatient days, b) STG nursing home inpatient days.

Hospital Outpatient Visits
We use two measures of *per-capita* hospital outpatient visits: a) total hospital outpatient visits,[124] b) clinic/referred (outpatient) visits.

F Additional Tables

Table F1: Descriptive Statistics: Random variables used for the Health Utilization Regressions (Health Utilization measures are in Per-Capita Terms)

	(1) Observations	(2) Mean	(3) Standard Deviation
Hospital Utilization rate (p.c.)	18474	0.3511	0.253
Total Hosp. Inpatient days (p.c.)	18474	0.7345	1.262
STG Hosp. Inpat. days (p.c.)	18474	0.6157	0.989
STNG-LT Hosp. Inpat. days (p.c.)	18474	0.1188	0.764
ST Community Hosp. Inp. days (p.c.)	18474	0.6130	0.988
ST Rehabilitation Inp. days (p.c.)	18474	0.0050	0.034
LT Rehabilitation Inp. days (p.c.)	18474	0.0002	0.006
Total Nursing Home Inp. days (p.c.)	18474	0.2390	0.817
STG Nursing Home Inp. days (p.c.)	18474	0.2279	0.801
Total Nursing Home Admissions (p.c.)	18474	0.0012	0.003
STG Nursing Home Admissions (p.c.)	18474	0.0011	0.003
STNG-LT Nursing Home Admissions (p.c.)	18474	0.0001	0.001
LT Total Outpatient visits (p.c.)	18474	0.0201	0.277
LT Clinic Outpatient visits (p.c.)	18474	0.0603	0.379
Insurance coverage (fraction)	18474	0.8185	0.059
White-non-Hispanic (fraction)	18474	0.7916	0.196
White-Hispanic (fraction)	18474	0.0717	0.125
Black-non-Hispanic (fraction)	18474	0.0866	0.143
Black-Hispanic (fraction)	18474	0.0023	0.004
Asian-non-Hispanic (fraction)	18474	0.0107	0.023
Asian-Hispanic (fraction)	18474	0.0004	0.001
Other-non-Hispanic (fraction)	18474	0.0322	0.080
Other-Hispanic (fraction)	18474	0.0046	0.006
Male (fraction)	18474	0.4993	0.022
Population 15-19 years old (fraction)	18474	0.0718	0.011
Population 20-24 years old (fraction)	18474	0.0615	0.025
Population 25-29 years old (fraction)	18474	0.0573	0.013
Population >64 years old (fraction)	18474	0.1545	0.041
$Self\ Employment_{it-1}$	18474	0.2217	0.080
$Self\ Employment_{it}^{Neighbor}$	18474	0.2265	0.084

Notes: p.c. denotes that the variable is in per-capita terms.

[124] The total hospital outpatient visits is the sum of three types of outpatient visits: hospital (outpatient) visits, clinic/referred visits, and ER (Emergency Room) outpatient visits.

Table F2: First-stage Regression (Dependent Variable: Insurance coverage)

	(1) Dep. Var.: Insurance Coverage
$Self\ Employment_{it-1}$	-0.111***
	(0.007)
$Self\ Employment_{it}^{Neighbor}$	-0.015***
	(0.006)
Observations	18474
R-squared	0.290
	F-test of excluded instruments
F(2,3078)	124.06
P-value	0.000
YEAR FE	YES
STATE-TIME DUMMIES	YES
CONTROLS†	YES

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis.
†: The list of controls at the county level include: eight race dummies, gender, four dummies for the age distribution, population density.

Table F3: First-stage Regression

	(1) Dep. Var.: Insurance Coverage
$Self\ Employment_{it+2}$	0.083***
	(0.020)
Observations	16772
Adj. R-squared‡	0.886
	F-test of excluded instruments
F(1,2939)	17.04
P-value	0.000
YEAR FE	YES
COUNTY FE	YES
CONTROLS†	YES

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis. †: The list of controls at the county level include the full set of controls: see the notes of Table F5.
‡: The Adjusted R-squared includes the effect of the County Fixed Effects, and is found using the command areg in Stata 12.

Table F4: First-stage Regression

	(1) Dep. Var.: Insurance Coverage
Female obesity (%)	-0.0014***
	(0.0003)
Male obesity (%)	0.0009***
	(0.0004)
Observations	16754
Adj. R-squared‡	0.886
	F-test of excluded instruments
F(2,2936)	14.31
P-value	0.000
YEAR FE	YES
COUNTY FE	YES
CONTROLS†	YES

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis. †: The list of controls at the county level include the full set of controls: see the notes of Table F5.
‡: The Adjusted R-squared includes the effect of the County Fixed Effects, and is found using the command areg in Stata 12.

Table F5: FE-IV Regressions: Second stage results (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
Ins. Coverage	27.188	11.224**	-10.630	1.376	0.162	0.449	6.676	0.633
	(27.927)	(4.524)	(13.778)	(1.096)	(0.299)	(0.955)	(6.331)	(1.271)
Hausman test (Coverage)								
Test statistic: $\chi^2_1 =$	0.430	6.211	0.557	1.732	0.276	0.280	1.304	0.550
p-value	0.512	***0.013***	0.456	0.188	0.599	0.597	0.254	0.458
Hansen J statistic§: $\chi^2_1 =$	0.296	0.007	0.056	0.593	0.046	0.002	0.500	0.054
p-value	0.586	0.935	0.812	0.441	0.830	0.963	0.479	0.816
Observations	16754	16754	16754	16754	16754	16754	16754	16754
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: *** $p < 0.01$, ** $p < 0.05$ * $p < 0.10$. Clustered Standard errors at the county level in parenthesis. The number of reported crimes by county are expressed per 1000 individuals and the insurance coverage is a fraction ranging from 0.505 to 0.967 at the county level. The unit of analysis is the county.
§: The instruments are: 1) the percentage of obese females (by county) and 2) the percentage of obese males (by county).
†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table F6: First-stage Regression

	(1)
	Dep. Var.: Insurance Coverage
$Self\ Employment_{it+2}$	0.0843***
	(0.0200)
$Diabetes\ Incidence_{it-1}$	-0.0012***
	(0.0003)
Observations	16772
Adj. R-squared‡	0.887
	F-test of excluded instruments
F(2,2939)	18.68
P-value	0.000
YEAR FE	YES
COUNTY FE	YES
CONTROLS†	YES

Notes: *** $p < 0.01$, ** $p < 0.05$ * $p < 0.10$. Clustered Standard errors at the county level in parenthesis. †: The list of controls at the county level include the full set of controls: see the notes of Table F5.
‡: The Adjusted R-squared includes the effect of the County Fixed Effects, and is found using the command areg in Stata 12.

Table F7: Reduced-Form Regressions (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
$Self\ Employment_{it+2}$	4.681*	0.958*	2.263	0.109	-0.059	-0.049	0.135	-0.148
	(2.456)	(0.491)	(1.922)	(0.217)	(0.048)	(0.125)	(0.669)	(0.106)
$Diabetes\ Incidence_{it-1}$	-0.073	-0.005	-0.004	0.003	3.41e-06	0.004**	-0.012	-0.0008
	(0.045)	(0.007)	(0.021)	(0.002)	(0.0005)	(0.002)	(0.012)	(0.0021)
Observations	16772	16772	16772	16772	16772	16772	16772	16772
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: *** $p < 0.01$, ** $p < 0.05$ * $p < 0.10$. Clustered Standard errors at the county level in parenthesis.
†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table F8: First-stage Regression

	(1)
	Dep. Var.: Insurance Coverage
$Self\ Employment_{it+2}$	0.0834***
	(0.0196)
$Diabetes\ Incidence_{it-1}$	-0.0011***
	(0.0003)
$Female\ Obesity_{it}$	-0.0011***
	(0.0003)
$Male\ Obesity_{it}$	0.0011***
	(0.0003)
$Female\ Obesity_{it-3}$	-0.0018***
	(0.0003)
Observations	16754
Adj. R-squared‡	0.888
	F-test of excluded instruments
F(5,2939)	20.39
P-value	0.000
YEAR FE	YES
COUNTY FE	YES
CONTROLS†	YES

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis. †: The list of controls at the county level include the full set of controls: see the notes of Table F5.
‡: The Adjusted R-squared includes the effect of the County Fixed Effects, and is found using the command areg in Stata 12.

Table F9: Reduced-Form Regressions (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
$Self\ Employment_{it+2}$	4.704*	0.945*	2.300	0.103	-0.059	-0.051	0.116	-0.148
	(2.451)	(0.491)	(1.924)	(0.219)	(0.048)	(0.124)	(0.668)	(0.106)
$Diabetes\ Incidence_{it-1}$	-0.072	-0.005	-0.007	0.003*	0.00004	0.004**	-0.012	-0.0008
	(0.045)	(0.006)	(0.021)	(0.002)	(0.0005)	(0.002)	(0.011)	(0.002)
$Female\ Obesity_{it}$	-0.013	-0.014*	0.015	-0.003	-0.0002	-0.0007	-0.013	-0.0006
	(0.049)	(0.008)	(0.024)	(0.002)	(0.0005)	(0.001)	(0.011)	(0.002)
$Male\ Obesity_{it}$	0.055	0.012	-0.005	-0.0006	0.0002	-0.00002	0.0003	0.001
	(0.048)	(0.009)	(0.024)	(0.003)	(0.0006)	(0.002)	(0.012)	(0.002)
$Female\ Obesity_{it-3}$	-0.062	-0.004	0.027	0.001	-0.0004	-0.002	0.011	-0.0008
	(0.047)	(0.008)	(0.023)	(0.003)	(0.0005)	(0.002)	(0.013)	(0.002)
Observations	16754	16754	16754	16754	16754	16754	16754	16754
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis.
†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table F10: FD (First Difference) Estimator: Breakdown by type of Crime (Crimes per 1000 individuals)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Property Crimes per 1000 individuals				Violent Crimes per 1000 individuals			
	Larceny Theft	Car Theft	Burglary	Arson	Murder	Rape	Aggravated Assaults	Robbery
Ins. Coverage	6.402***	0.355	-0.164	-0.297	0.042	-0.056	-0.333	-0.300***
	(1.779)	(0.353)	(1.176)	(0.217)	(0.054)	(0.153)	(0.527)	(0.103)
Observations	13433	13433	13433	13433	13433	13433	13433	13433
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES
COUNTY FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls†	ALL	ALL	ALL	ALL	ALL	ALL	ALL	ALL

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis.
†: The list of controls at the county level include: unemployment rate, population density, poverty rate, real personal income, the amount of per-capita social security recipients, 8 dummies for race, total police forces in year $(t-1)$, 3 variables for the age distribution; the percentage of small firms (< 500 employees) on the total number of firms, the number of per-capita arrests for sale and possession of drugs, the number of per-capita arrests for gambling. More details are presented in Section 3.

Table F11: First-stage Regression

	(1) Dep. Var.: Insurance Coverage
$\Delta Female\ obesity_{it}$	-0.0009***
	(0.0002)
$\Delta Home\ Health\ Agencies_{it}$	-0.0001***
	(0.00001)
$\Delta LT\ Retarded\ Hosp.\ Days_{it}$	1.32e-06***
	(3.74e-07)
$Self\ Employed_{it}$	0.0187***
	(0.0021)
Observations	13418
	F-test of excluded instruments
F(4,2917)	36.21
P-value	0.000
YEAR FE	YES
COUNTY FE	YES
CONTROLS†	YES

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis. †: The list of controls at the county level include the full set of controls: see the notes of Table F5.

Table F12: First-stage Regression

	(1) Dep. Var.: Insurance Coverage
$Self\ Employed_{it+2}$	0.0805***
	(0.0199)
$Female\ Obesity_{it}$	-0.0012***
	(0.0003)
Observations	16754
Adj. R-squared‡	0.887
	F-test of excluded instruments
F(2,2936)	17.39
P-value	0.000
YEAR FE	YES
COUNTY FE	YES
CONTROLS†	YES

Notes: ***$p < 0.01$, **$p < 0.05$ *$p < 0.10$. Clustered Standard errors at the county level in parenthesis. †: The list of controls at the county level include the full set of controls: see the notes of Table F5.
‡: The Adjusted R-squared includes the effect of the County Fixed Effects, and is found using the command areg in Stata 12.

Table F13: Reduced Form Regression

	(1)
	Dep. Var.: Pedestrian Fatality Rate
$Self\ Employed_{it+2}$	-0.242*
	(0.128)
$Female\ Obesity_{it}$	0.004*
	(0.002)
Observations	16754
YEAR FE	YES
COUNTY FE	YES
CONTROLS†	YES

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$ $^{*}p < 0.10$. Clustered Standard errors at the county level in parenthesis. †: The list of controls at the county level include the full set of controls: see the notes of Table F5.

G The determinants of obesity

Factors Contributing to Excess Caloric Intake:

- Increased consumption of sugar-sweetened beverages (Duffey & Popkin, 2007; Nielsen et al., 2002)
- Increased snacking (Duffey & Popkin, 2011; Jahns et al., 2001; Zizza et al., 2001)
- Larger portion sizes (Piernas & Popkin, 2011; Young & Nestle, 2002; Young & Nestle, 2003)
- Higher calorie-density of foods (Kant & Graubard, 2006)
- More meals consumed or purchased away from home (Kant & Graubard, 2004)
- More exposure to advertising that encourages food consumption and promotes unhealthy foods (French et al., 2001; Powell et al., 2011)
- Value-sizing of less nutritious foods (e.g., value meals at fast food outlets) (French et al., 2001; French, 2005)

Factors Contributing to Inadequate Amounts of Physical Activity

- Labor-saving technological advances (e.g., computers) (Sallis & Glanz, 2009)
- Increased media use (e.g., television, video games) (French et al., 2001)
- Automobile-oriented communities and reliance on motorized transportation (Sallis & Glanz, 2009)
- Limited access to safe, convenient recreation facilities or walking areas (French et al., 2001)
- Limited opportunities for activity during the workday (French et al., 2001)
- Limited time for daily physical education and recess in schools (Lee et al., 2007)