Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2010/05 35 pàg
Working Paper 2010/05 35 pag.

"DISCRETE DISTRIBUTIONS WHEN MODELING THE DISABILITY SEVERITY SCORE OF MOTOR VICTIMS"

Jean-Philippe Boucher and Miguel Santolino

RISC-IREA Universitat de Barcelona

Institut de Recerca en Economia Aplicada | *Research Institute of Applied Economics*

Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

WEBSITE: www.ub.edu/irea/ • CONTACT: irea@ub.edu

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

Abstract

Many European states apply score systems to evaluate the disability severity of non-fatal motor victims under the law of third-party liability. The score is a non-negative integer with an upper bound at 100 that increases with severity. It may be automatically converted into financial terms and thus also reflects the compensation cost for disability. In this paper, discrete regression models are applied to analyze the factors that influence the disability severity score of victims. Standard and zero-altered regression models are compared from two perspectives: an interpretation of the data generating process and the level of statistical fit. The results have implications for traffic safety policy decisions aimed at reducing accident severity. An application using data from Spain is provided.

.

JEL classification: -

Keywords**:** Hurdle discrete data models, zero-inflated distribution, generalized method of moments, personal injuries, disability rating scale.

Jean-Philippe Boucher. Département de Mathématiques. Université du Québec à Montréal. Québec, Canada
Miguel Santolino is Teaching Assistant at Universitat de Barcelona. Department of Econometrics,. E-mail: msantolino@ub.edu

Acknowledgements:

Jean-Philippe Boucher gratefully acknowledges the financial support of the Conseil de recherches en sciences naturelles et en génie du Canada. Miguel Santolino wishes to acknowledge the support of the Spanish Ministry of Science and Innovation and FEDER grant ECO2008-01223/ECON. The authors are grateful to the director of the Risk in Finance and Insurance Group at the University of Barcelona, Montserrat Guillén, for comments and suggestions.

1. Introduction

The analysis of factors that influence the injury severity of victims involved in motor accidents is a major issue for a number of areas of traffic safety. In many European countries score systems are used in the field of third-party liability to evaluate the disability severity of non-fatal victims involved in traffic accidents. These score systems represent the disability severity by a single numerical value or score. The score indicates the disability severity associated with all the injuries sustained by the victim, and represents the percentage of difficulty experienced when performing the customary movements and actions of everyday living. The score value ranges between zero, for a victim without disability, to a theoretical maximum which is usually set at one hundred.

The application of disability score systems is especially useful when assessing the financial compensation associated with motor injuries. The monetary valuation of the non-financial prejudice sustained by the victim for the injuries is legally fixed according to the severity score[1]. Therefore, the severity score not only reflects the disability severity but is also an indicator of the monetary cost of the non-financial damage suffered by the victim. Consequently, it is important to use adequate statistical models when analyzing factors that influence the disability score in order to guide traffic safety policy decisions aimed at reducing the severity and cost of disabilities. States with score systems include Italy, Portugal, Belgium, France and Spain. In addition, a project is underway in the European Union to harmonize disability assessment practices by applying a European disability rating scale (EC, 2003).

The present study applies discrete regression techniques to the modeling of disability severity scores of motor victims. More specifically it compares the use of the Poisson, Geometric and Negative Binomial models for the analysis of disability severity data. As there may be important differences between injured victims who did not suffer any disability (a zero score) and those left with permanent disabilities, zero-altered distributions, such as the zero-inflated and the hurdle models, are also considered. These zero-altered distributions assume that disability severity data are generated by a dual-state process (Lord *et al*., 2005; 2007; Ullah *et al*., 2009). Regression models are compared from two perspectives, namely the interpretation of the underlying data generating process and the level of statistical fit. An application to a Spanish database is provided, in which the factors that influence the degree of disability severity are examined.

A variety of statistical models have been applied in the literature to investigate the effects of road conditions, driver attributes or vehicle characteristics on injury severity. Traditionally, the severity outcome of traffic victims has been defined as a qualitative variable that consists of categories which reflect the underlying victim severity. The response categories are recorded on ordinal scales from lower to higher severity levels (Yamamoto *et al.*, 2008; Huang *et al*., 2008; Delen *et al*., 2006; Jung *et al*., 2010). The distance between adjoining categories is difficult to establish because there are many factors and circumstances that define the physical condition of an accident victim. As a result, in the regression context, statistical methods for modeling categorical variables have been the most extensively applied techniques in the analysis of motor injury severity.

One of the most popular techniques for modeling injury severity is the ordered multiple-choice model (O'Donnell and Connor, 1996; Abdel-Aty and Keller, 2005; Kockelman and Kweon, 2002; Khattak and Rocha, 2003, Wang and Kockelman, 2005), which takes into account the order of response categories. In some cases ordered multiple-choice structures may lack sufficient flexibility, due, for example, to constraints related to opposite marginal effects on the two extreme categories of the response variable. Recent publications have extended these models, including random effects to relax some of their restrictions (Eluru *et al*., 2008; Srinivassan, 2002). An alternative approach is to apply multiple-choice model structures in which the order of categories is ignored, as in the case of multinomial and nested logit/probit models (Eluru and Bhat, 2007; Milton *et al*., 2008; Savolainen and Mannering, 2007, Malyshkina and Mannering, 2009).

[1] The law in some countries, such as Spain or Italy, also considers the age of the victim when calculating the monetary valuation of the disability.

Unlike the qualitative modeling approach the distance between disability severity scores is constant and, consequently, statistical techniques for modeling quantitative injury-severity data can be applied. Modeling the severity score is more flexible than modeling qualitative levels. For instance, let us suppose that a mild disability is associated with severity scores below 15 points, and a severe disability with 15 points or more. In the (ordered) multiple-choice modeling structure a victim with a score of 1 point and another with a score of 14 are both classified as mild disabilities. If these two victims are incorrectly predicted by the model as severely disabled, the classification error made is the same in both cases. Obviously, these two victims do not have the same degree of disability severity and the error is larger for the first one. An obvious constraint of qualitative structures is that the performance of the regression models may be influenced by the definition of severity levels considered in the response categories.

In our context the score for disability severity is clearly defined as the percentage of disability, and hence it is a more precise measure of the degree of severity. Therefore, methods that model the disability severity score are an improvement because they are not affected by any partial definition of severity categories. In addition, the score estimate may be directly expressed in financial terms since the compensation for the non-financial prejudice is determined by the size of the disability severity score. By contrast, a summary measure of the cost should be used when severity categories are considered, an example being the mean cost of each category.

The structure of the paper is as follows. Section 2 presents basic discrete distributions that can be used to model the severity score, as well as their zero-altered extended models. The reasons for using - and interpretations of - the distributions are explained. In section 3, empirical applications using a Spanish database are presented and statistical comparisons between models are described. A special test concerning the difference between zero-inflated and hurdle models is also described. This is followed by discussion and interpretation of the results. Concluding remarks are summarized in Section 4.

2. Modeling the disability severity

There are many discrete probability distributions that can be used to model the severity score of an injured motor victim. Obviously, it is important to verify the statistical goodness-of-fit of these distributions with real data. However, looking at the properties and characteristics of each distribution is also useful in terms of understanding how a specific distribution may give appealing interpretations of the data generating process.

Let the response variable take the value y_i, which is the severity score for the permanent disability sustained by the motor victim i resulting from a traffic accident: $i = 1, \ldots, n$. Because the score is limited to values from 0 to 100, then the probability function of the severity score Z_i is $\Pr(Z_i = y_i) = \Pr(y_i) / \Pr(Y_i < 100)$, where Y_i is a discrete variable. Probability functions of Y_i are described below.

2.1. Basic distributions

The starting point for the modeling of a random variable that takes positive discrete values is commonly the Poisson distribution. If we suppose the score is Poisson distributed, then the probability function of Y_i would be:

$$\Pr[Y_i = y_i] = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}, \quad y_i = 0, 1, 2, \ldots, \tag{1}$$

where the $(k \times 1)$ vector of explanatory variables x_i is included in the model with the mean parameter $\lambda_i = \exp(x_i' \beta)$, where β is the $(k \times 1)$ vector of parameters to be estimated.

The Poisson distribution is known to be the *law of rare events* or *law of small numbers*, since it can be shown to be the limit of a binomial distribution with the number of attempts going to infinity and the probability of success tending to 0. This interpretation of the Poisson distribution is rather difficult to implement in the context of the severity score. Nevertheless, it is worth noting that the Poisson distribution is commonly used for discrete data and is the basis for some extensions.

A more intuitive distribution can, however, be used to model the severity score. If we consider the final severity score as an indicator variable that takes higher values if the injury is more serious, then the Geometric distribution would seem to be a natural choice. This distribution is known to model the *number of successes before a single failure*.

Applied to the severity score the justification of the Geometric distribution is as follows. To obtain a specific severity score y_i, the i-*th* victim must have all the symptoms associated with values $1,2,\ldots,y_i$, which we call *successes*, without having all the symptoms of score y_i+1, which represents a *failure*. Using this interpretation the score probability can be expressed as:

$$\Pr[Y_i = y_i] = p_i^{y_i}(1-p_i), \tag{2}$$

where covariates can be included in the model using a logit function, and thus $p_i = \exp(x_i'\beta)/(\exp(x_i'\beta)+1)$. In this case, it can be shown that $E[Y_i] = p_i$ and $Var[Y_i] = p_i(1-p_i)$.

Another alternative would be the Negative Binomial (NB) distribution. Although there are many ways to construct a NB distribution, one of the most convenient is to present it as a generalization of the Geometric distribution, in which it models the *number of successes before a specific quantity of failures*. Another approach consists in introducing a random heterogeneity term θ of mean 1 and variance α into the mean parameter of the Poisson distribution. If variable θ follows a gamma distribution and has the following density distribution:

$$f(\theta) = \frac{(1/\alpha)^{1/\alpha}}{\Gamma(1/\alpha)} \theta^{1/\alpha - 1} \exp(-\theta/\alpha),$$

then the mixture will result in a NB distribution with probability distribution:

$$\begin{aligned} \Pr[Y_i = y_i] &= \frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})} \left(\frac{\lambda_i}{\alpha^{-1}+\lambda_i}\right)^{y_i} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\lambda_i}\right)^{\alpha^{-1}} \\ &= \frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})} p_i^{y_i} (1-p_i)^{\alpha^{-1}} \end{aligned} \tag{3}$$

where the $\lambda_i = \exp(x_i^{'}\beta)$, α is a positive parameter and $\Gamma(\cdot)$ is the gamma function, defined by $\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt$ (for early discussions of NB distributions see Greenwood and Yule, 1920; for extensions designed to include exogenous data and an application to accident data, see Lawless, 1987; and Dionne and Vanasse, 1989). The second equality in (3) highlights the fact that the logit transformation of the regressors of the Geometric distribution is now generalized by $p_i = \left(\frac{\lambda_i}{\alpha^{-1}+\lambda_i}\right)$. It can then easily be proved that $E[Y_i] = \lambda_i$ and $Var[Y_i] = \lambda_i + \alpha\lambda_i^2$.

2.2. Zero-altered models

Motor victims with only temporary disabilities resulting from an accident have a score equal to zero, due to the non-permanent nature of their disability. Intuition would suggest that accidents involving permanent disabilities may not have the same characteristics as those involving only temporary disabilities, so zero-altered distributions should also be considered. This section analyzes zero-inflated and hurdle models based on the three basic distributions presented above (Poisson, Geometric and Negative Binomial).

2.2.1 Zero-inflated models

A distribution with extra zeros can show a good fit to data that exhibit a high frequency of zeros. The idea of zero-inflated (ZI) models is to use a finite mixture model combining an indicator distribution for the zero case with a basic discrete distribution. Consequently, this distribution will account for the excess of zeros. The density of this kind of model, with $0 < \phi_i < 1$, can be expressed as:

$$P(Y_i = y_i) = \begin{cases} \phi_i + (1-\phi_i)\Pr(K_i = 0) & \text{for } y_i = 0 \\ (1-\phi_i)\Pr(K_i = y_i) & \text{for } y_i = 1, 2, \ldots \end{cases} \tag{4}$$

where the random variable K_i follows a basic distribution as defined in section 2.1. In the limiting case, where $\phi_i \to 0$, the zero-inflated model corresponds to the distribution of K_i. Mullahy (1986) used this distribution with constant zero-inflation ϕ, while in our context the $(p \times 1)$ vector of regressors w_i is included such that $\phi_i = \exp(w_i'\gamma)/(\exp(w_i'\gamma)+1)$, and γ is the $(p \times 1)$ vector of parameters to be estimated.

The first two moments of the ZI distribution are $E[Y_i] = (1-\phi_i)E[K_i]$ and $Var[Y_i] = (1-\phi_i)E[K_i^2] - (1-\phi_i)^2 E[K_i]^2$. Because the difference between the ZI model and the basic distribution is the extra weight for $Y_i = 0$, it is easy to adjust the maximum likelihood equations of the basic distribution to find the parameters of the ZI model. See Lambert (1992) for an application of a zero-inflated Poisson model or Johnson et al. (1995) for an overall discussion of this simple way to account for the extra zeros.

2.2.2 Hurdle models

An alternative way of modifying a basic discrete distribution is to use it as part of a two-process distribution, there being one process below and another above the hurdle. The hurdle model was introduced by Cragg (1971), and subsequently reviewed by Mullahy (1986). The first process is a dichotomous distribution that differentiates victims with a zero score from those with a positive score. The second process determines the final severity score, which depends on a score greater than 0 having been reported. Distributions that use two processes may be driven by the same explanatory variables, although they will be interpreted according to the processes involved.

The first part of the model is a binary outcome model, while the second is a discrete distribution that takes the values in $\{1,2,3,\ldots\}$. Consequently, when modeling the second part, one must choose between a basic discrete distribution (truncated or shifted) and a discrete distribution with support domain on $\{1,2,3,\ldots\}$. Formally, without loss of generality, this hurdle model is expressed as follows. Let $f_{i,1}(\cdot)$ and $f_{i,2}(\cdot)$ be two probability mass functions with respective support $\{0,1\}$ and $\{0,1,\ldots\}$ and which depend on parameter vectors θ_1 and θ_2. The random variable Y_i obeys the hurdle distribution if:

$$P(Y_i = y) = \begin{cases} f_{i,1}(0) & \text{for } y = 0 \\ \frac{1-f_{i,1}(0)}{1-f_{i,2}(0)} f_{i,2}(y) = \Psi_i f_{i,2}(y) & \text{for } y = 1,2,\ldots \end{cases}, \tag{5}$$

where $\Psi_i = \frac{1-f_{i,1}(0)}{1-f_{i,2}(0)}$. If we assume $f_{i,2}(0) = 0$, then $\Psi_i = 1 - f_{i,1}(0)$. The expectation and variance of Y_i are:

$$E[Y_i] = \Psi_i E_{f_{i,2}(k)}[K_i],$$
$$Var[Y_i] = \Psi_i E_{f_{i,2}(k)}[K_i^2] - \Psi_i^2 E_{f_{i,2}(k)}[K_i]^2,$$

where $E_{f_{i,2}(k)}[K_i]$ and $E_{f_{i,2}(k)}[K_i^2]$ are the first two moments of distribution $f_{i,2}$. Thus, the distribution can be over- or underdispersed, depending on the parent processes $f_{i,1}$ and $f_{i,2}$. Many possible choices exist for the parent-processes $f_{i,1}$ and $f_{i,2}$, for example, nested models in which $f_{i,1}$ and $f_{i,2}$ come from the same distribution, such as the Poisson distribution (Mullahy, 1986), or the Negative Binomial (Pohlmeier and Ulrich, 1995) which is by far the most popular hurdle model (Winkelmann, 2003a). Non-nested models such as those offered by Grootendorst (1995), Gurmu (1998) and Winkelmann (2003a) can also be used. In our case, $f_{i,1}$ is Bernoulli distributed with parameter δ_i and $f_{i,2}$ follows a basic distribution as described in section (2.1). Rather than nest with basic discrete distributions, hurdle models overlap (Vuong, 1989) because they can be equivalent for certain parameter restrictions.

It is easy to estimate parameters by maximum likelihood because each process may be estimated separately. The log-likelihood function of a hurdle model is expressed as:

$$\ell = \sum_{i=1}^{n} \left[I_{(y_i=0)} \log(f_{i,1}(0)) + I_{(y_i>0)} \log(1-f_{i,1}(0)) \right] + \sum_{i=1}^{n} \left[I_{(y_i>0)} \log(f_{i,2}(y_i)/(1-f_{i,2}(0)) \right],$$

which is separable. Maximization can then be done separately for each part (zero case and positive values) with standard statistical software.

2.2.3 Model comparisons and interpretation

Zero-inflated and hurdle models can be expressed as a compound sum of two random variables (Boucher and Guillen, 2008), where *i* subscripts are removed to simplify annotation:

$$Y = \sum_{j=1}^{M} X_j, \tag{6}$$

where X_j is i.i.d., independent from M, and $Y = 0$ if $M = 0$. Under this construction, there are two possibilities:

- For the ZI distribution: $M \sim Bernoulli(\phi)$ with X_j taking values $0,1,2,3,\ldots$.
- For the hurdle distribution: $M \sim Bernoulli(\delta)$ with X_j taking only positive values $1,2,3,\ldots$.

Thus, one main difference between the two distributions is the way in which a zero is obtained. For the hurdle distribution a zero occurs only if $M=0$, while for the ZI model it occurs if $M=0$ or if $M=1$ and $X_1=0$. Obviously, and as shown by many authors (for example, Baughman, 2007), by taking a specific parameterization for δ, the hurdle distribution may be equivalent to the ZI distribution. In the context of the analysis of bodily injuries the differences between these two distributions can be usefully interpreted. Indeed, the characteristics of a zero severity score are interpreted differently depending on the construction of the compound sum. From an intuitive point of view, under the hurdle distribution hypothesis a zero severity score comes from different kinds of accidents than those which yield positive scores. By contrast, the ZI distribution supposes that the same kind of accident generates all severity scores. However, each accident has an extra probability (ϕ) of generating a score of 0.

The conditional moment (CM) test proposed by Santos-Silva and Windmeijer (2001) can be adapted to check if the hurdle specification is valid[2]. When the *separation hypothesis* assumption of the hurdle model is fulfilled the equality $E\big[Y-E[M]E[X]\big]=0$ is satisfied. This equality may be tested using the CM test described by Newey (1985) and Tauchen (1985), and further explained in Cameron and Trivedi (1998). The CM test is elaborated upon in the appendix.

3. Empirical application

3.1 Spanish database

The database consists of a random sample of $18,363$ non-fatal motor victims. The data were provided by one of the biggest motor insurance companies in Spain. All of the victims needed at least one day to recover from the injuries caused by the accident. The sample covers all provinces of Spain. Since an at-fault system is in place in Spain, drivers who were at fault in the accident are not entitled to compensation and, therefore, the severity of their injuries was not evaluated. Consequently, at-fault drivers were not included in the database. All victims were compensated for their injuries in 2007, although the accident may have occurred before that year.

The dependent variable to model is *score*, which indicates the victim's score for the permanent disability resulting from the accident. The score reflects the degree of difficulty the victim has in performing activities of the daily living, measured from zero to an upper limit set at 100. A zero score is assigned to victims who are fully recovered from their injuries and, thus, without permanent disability. The score is determined according to a legislative score system. The score system consists of a medical scale that describes the possible injuries resulting from traffic accidents and which provides maximum/minimum point scores for each one. The final score is awarded by judicial decision, or agreed upon between parties (insurer and victim), based on medical reports provided by parties and a forensic examination. Medical reports are made by medical specialists who examined (and evaluated) the injuries of the victim in accordance with the score system. Therefore, the score system is a two-step evaluation process. First, it must be decided whether the victim suffered any of the injuries defined in the scale. Subsequently, a score reflecting the disability severity caused by injuries must be provided as a sum of points with an upper limit at 100 (for more details on the Spanish system, see Ayuso and Santolino, 2007; Ayuso *et al.*, 2010).

The distribution of the disability severity score for our data is presented in Figure 1. The sample's mean score is 3.93 points, the standard deviation is 6.51 and the median value is 2. Note that most of the observations are concentrated on the part associated with low score values, between one and three points approximately. The distribution seems to be zero-deflated. Most of the sample victims suffered permanent disability resulting from the

[2] In the paper by Santos-Silva and Windmeijer (2001) the same test is used to verify what they call *the multiple illness spells hypothesis* in order to distinguish the hurdle distribution from a special case of the Negative Binomial distribution.

accident. Fewer than 7% suffered only temporary injuries (a zero score). One possible explanation for the presence of many non-zero scores is that the Spanish score system recognizes whiplash as an injury which causes permanent disability. Whiplash injuries are very common in motor crashes. In this sample, 34% of the injuries sustained by victims were associated with whiplash.

Figure 1. Histogram for the disability severity score

[INSERT FIGURE 1]

The explanatory variables in the regression model and main statistics are provided in Table 1. Individual attributes of the victim are recorded in the variables *gend* and *age.* Three additional variables relate to the position of the victim inside or outside the vehicle. These variables identify whether the victim was the driver, a passenger or either a pedestrian or cyclist. Three variables are associated with the period that the victim needed to recover from the injuries. The temporary period that the victim was in hospital is recorded in the variable *hrd.* The days that the victim was out of hospital but was unable to work are recorded in variable *ird.* Finally, the variable *nird* indicates the number of out-of-hospital recovery days on which the victim was able to work but still needed some therapy, such as going to a rehabilitation centre. To avoid endogeneity, values collected in these three variables were based on the forensic examination rather than the final values stated by courts or agreed between parties.

Table 1. Description of variables and some statistics

Variable	Description	Mean	Std.Dev.
score	Severity score of permanent disability.	3.930	6.510
gend	1 if the injured victim is male; 0 otherwise.	0.451	0.497
age	Age of the victim.	38.251	17.029
driv	1 if the injured victim was the driver; 0 otherwise.	0.482	0.500
pas	1 if the injured victim was the passenger; 0 otherwise.	0.383	0.486
pedcy	1 if the injured victim was either a pedestrian or a cyclist; 0 otherwise.	0.134	0.341
hrd	Number of recovery days in hospital (in log scale)*.	0.341	0.891
ird	Number of out-of-hospital recovery days with inability to work (in log scale)*.	3.900	1.254
nird	Number of out-of-hospital recovery days without inability to work (in log scale)*.	2.026	1.972

* One point was added on the value of these variables before conversion into logarithms.

3.2. Parameter estimates and model comparisons

In this section, parameter estimates are shown and an interpretation of the results is provided for all the models considered in this paper. For some specific parameter restrictions, the Negative Binomial is equivalent to the Poisson or the Geometric distributions. Classical hypothesis tests can be performed to accept or reject nested models, with some precautions being taken when the parameter restriction corresponds to the boundary of the parameter space. The interested reader can find a description of some specification tests for choosing between Poisson and Negative Binomial distributions in basic, zero-inflated or hurdle models in Boucher *et al*. (2007; 2009). Application of these tests to our score data for the α parameter of the NB distribution leads to the rejection of Poisson and Geometric distributions in favor of the Negative Binomial for basic (both *p-values* less than 0.001), zero-inflated (both *p-values* less than 0.001), and hurdle (both *p-values* less than 0.001) constructions.

However, the question of how a severity score of zero is generated remains undecided. There are three candidates for modeling the severity score: the basic, the zero-inflated and the hurdle Negative Binomial regressions. Estimated coefficients of these models are provided in Table 2. Vector a consists of the parameters of the dichotomous process, while

β comprises the parameters of the NB process.

The Akaike and Bayesian information criteria (AIC and BIC, respectively) clearly give an advantage to the hurdle NB model. To analyze whether the observed differences in the log-likelihood and the information criterion are statistically significant, a test based on the difference in the log-likelihoods can be performed. Indeed, for independent observations a log-likelihood ratio test for non-nested models developed by Vuong (1989) can be used to see whether the hurdle NB model is statistically better than the zero-inflated NB and the NB model. Applied to the present data, the Vuong test shows that the information criterion of the hurdle NB model is statistically different from the other models, with *p-values* less than 0.001 for each test. Therefore, from a statistical viewpoint the hurdle NB model offers the best fit.

Table 2. Parameter estimates (standard errors) of the Negative Binomial (NB), the zero-inflated Negative Binomial (ZI-NB) and the hurdle Negative Binomial (Hurdle-NB) models

	NB		ZI-NB		Hurdle-NB	
a_0	-	-	-0.802	(0.687)	1.217	(0.156)
a_{gend}	-	-	*	*	0.644	(0.068)
a_{age}	-	-	-0.251	(0.030)	-0.049	(0.002)
a_{pas}	-	-	*	*	0.366	(0.074)
a_{pedcy}	-	-	1.999	(0.579)	1.286	(0.092)
a_{hrd}	-	-	*	*	*	*
a_{ird}	-	-	*	*	-0.654	(0.026)
a_{nird}	-	-	*	*	-0.312	(0.021)
β_0	-0.932	(0.035)	-0.901	(0.035)	-2.088	(0.056)
β_{gend}	-0.043	(0.013)	-0.041	(0.013)	*	*
β_{age}	0.008	(0.000)	0.008	(0.000)	0.009	(0.001)
β_{pas}	-0.040	(0.013)	-0.034	(0.014)	*	*
β_{pedcy}	*	*	0.047	(0.019)	0.180	(0.028)
β_{hrd}	0.443	(0.006)	0.445	(0.006)	0.549	(0.010)
β_{ird}	0.372	(0.006)	0.370	(0.006)	0.519	(0.010)
β_{nird}	0.045	(0.003)	0.045	(0.003)	0.053	(0.005)
alpha	0.345	(0.006)	0.336	(0.006)	0.887	(0.017)
Loglikelihood	-39,233.84		-39,166.13		-36,770.41	
AIC	78,484		78,357		73,569	
BIC	78,546		78,451		73,678	

n = 18,363

* Coefficients were not significant at the 5% level. These variables were not included as regressors.

The CM test was computed to analyze whether the hurdle specification was accepted. The test rejected the null hypothesis, with a test value of 91.71 for a Chi-square with 8 degrees of freedom, and a *p-value* less than 0.001. However, this does not necessarily mean that it is unreasonable to assume that a zero score comes from a single process, because the test might reflect another kind of misspecification. Moreover, it is well-known that CM tests are very powerful when used with many observations (here, $n = 18,363$), meaning that the null hypothesis is usually rejected.

Another explanation for rejecting the null hypothesis is the possibility that common unidentified individual characteristics, i.e. heterogeneity, affect both processes of the hurdle model. In this case the separation hypothesis would be rejected. Winkelmann (2003b) developed a hurdle model with correlation between the two heterogeneity processes. Application of this model to our score data does not give a result where the correlation parameter is statistically different from zero.

3.2.1 The underlying data generating process

In addition to statistical tests a discussion about the underlying data generating process may prove useful for the model selection analysis. It has been shown that two-process models take into account that zero scores have different characteristics from positive scores, which means that accidents involving permanent injuries have different features from those involving only temporary injuries. This would seem to be a reasonable assumption and hence regression models involving two processes may be adequate techniques for describing the data generating process.

Comparison of two-process models revealed that the hurdle model performed better than the ZI model in our application. This result was expected because the data plotted in Figure 1 show a lower percentage of victims with only temporary disabilities than would be expected with the basic distribution. ZI models add an extra weight to the zero score, and thus they are more adequate for overdispersed data. Hurdle models are also more suitable to explain the data generating process. Under the hurdle assumption, zeros only come from accidents with temporary-injury features. By contrast, zero-inflated models assume that victims with very mild permanent injuries are also associated with scores equal to zero. The Spanish score system always allocates positive scores for all permanent injuries. This means that victims who suffered any of the injuries included in the medical scale cannot have a zero score for disability. Consequently, only one source of zeros seems a more satisfactory interpretation. Based on these arguments it can be concluded that the hurdle model provides the best description of the score process, even though the CM test rejected it.

3.2.2 Results and discussion

The rest of the paper focuses on analyzing the parameter estimates of the hurdle model. An advantage of modeling disability by means of a hurdle regression is that it allows the factors that influence the probability of suffering a permanent disability (below the hurdle) to be analyzed separately from those that affect the degree of severity (above the hurdle). Note that males have a lower probability of suffering a permanent disability than do females ($\hat{a}_{gend} = 0.644$). However, conditioned on having permanent disability, gender does not show explanatory capacity as regards the level of severity sustained. In the literature it is broadly accepted that females involved in motor accidents suffer more serious injuries than do males (Evans, 2001; Kockelman and Kweon, 2002). This would be in accordance with our findings in the sense that females are more likely to suffer a permanent disability, although this does not hold for the degree of this disability.

The victim's age positively influences the frequency and severity of disabilities, as shown by the sign of the estimates of a_{age} and β_{age}. This means that older victims are more likely to suffer a permanent disability and, conditioned on having a permanent disability, aging is also related to a higher severity score. This result is consistent with the previous literature (O'Donnell and Connor, 1996; Delen et al., 2006). Other authors have suggested that young and old victims have more serious injuries, especially when they are the drivers. This result has been associated with the fact that young drivers are prone to more reckless driving and old drivers to slower reaction times (Kockelman and Kweon, 2002; Huang et al., 2008).

Regarding the victim's position inside the vehicle, the base category is the driver. According to the results, pedestrians or cyclists are less likely to suffer a permanent disability than are drivers. However, if a permanent disability is sustained, pedestrians and cyclists experience higher levels of severity. Our intuition was that victims such as pedestrians and cyclists would always sustain more serious damages, yet the empirical results show the opposite. A potential explanation for this is that although crashes involving pedestrians and cyclists are likely to cause injuries to these victims, such injuries normally have only temporary consequences.

The variable related to passengers shows a positive coefficient in the binary outcome model. Therefore, passengers are less likely to have a permanent disability than are drivers. This could be because drivers display safer driving behavior when they are accompanied and, therefore, passengers are more likely to be involved in accidents causing only temporary injuries. Some studies have shown that the presence of passengers prevents risky driving behavior among drivers, for instance, by warning the driver (Rueda-Domingoa *et al*., 2004; Lee and Abdel-Aty, 2008). Other studies, however, have shown that passengers are associated with more serious injuries than are drivers (O'Donnell and Connor, 1996; Ulfarsson and Mannering, 2004). According to our results, passengers do not show a different degree of disability severity from drivers when an accident involving permanent disabilities occurs.

Finally, let us consider the relationship between the recovery period required by victims and the permanent disability sustained. The recovery period relates to the time needed for the injuries to stabilize or be fully resolved. Note that five of the six variables related to the temporary recovery period show significant coefficients, with the expected sign. No collinearity problems were detected. The longer the out-of-hospital recovery period, the greater the likelihood of a permanent disability. In addition, the length of the recovery period is also positively associated with the severity score for permanent disability, including the time spent in hospital. The results are also consistent with the type of recovery. The number of recovery days in hospital shows the greatest influence on the expected severity, followed by the out-of-hospital days with inability to work and, finally, the out-of-hospital days without inability to work. The effect on the disability probability according to type of recovery is also as expected.

4. Conclusion

Well-developed statistical regression models have been applied in order to analyze severity injury data, even though published studies have, to date, modeled the injury severity as a qualitative dependent variable. These modeling approaches are strongly dependent on the definition used for the categories of injury severity and may not be flexible enough (for a discussion about constraints of ordinal qualitative approaches see, for instance, Ulfarsson and Mannering, 2004). The present study contributes to the field by directly applying regression techniques for quantitative dependent variables to the analysis of disability severity score data. The most widely-used basic and zero-altered discrete distributions and regression models are compared from the point of view of the underlying data generating process and the level of fit.

An application to a Spanish motor disability database is provided in which the hurdle-Negative Binomial regression model was the preferred method for this dual approach. The hurdle structure based on two processes offers the most plausible interpretation of the process that generates the Spanish data. An advantage of hurdle models is that the zero score process is modeled separately, and may thus be analyzed independently. The hurdle part models the probability of suffering permanent injuries, while the discrete distribution refers to the disability severity of these injuries. Statistical tests supported the selected hurdle model as the construction that offers the best fit to the data. The separation hypothesis of the hurdle model was explored, although the model specification was rejected; this was most likely due to the power of the test, or perhaps to the fact that some confusion at zero cannot be captured by the model complexity. For instance, some permanent disability, which could include psychological damage, may not qualify for a positive score.

Among other results the victim’s gender showed an influence on the probability of suffering permanent disabilities but not on the degree of disability severity caused by these injuries. Another interesting finding is that the length of the time required by the victim to recover from the temporary injuries had strong explanatory power as regards the probability of suffering a permanent disability and the degree of severity of this disability. This result sheds light on the relationship between temporary and permanent injuries, which is important for medical specialists, among other key players in this context. Finally, it is worth highlighting that the score estimation can be automatically converted into financial terms, indicating the compensation cost for the non-financial prejudice resulting from permanent injuries. Therefore, the analysis of factors that affect the degree of disability may guide policy planners in tackling the motor injury problem not only from the viewpoint of the severity of physical disabilities caused by accidents but also in terms of the financial consequences borne by society.

Appendix

By noting that the first process of the hurdle model involves only the random variable M, while the second process uses the variable X_j, the conditional moment test is constructed on the following statements. If the hurdle hypothesis is true, the following moment conditions for the first and second processes hold:

$$E[I_{(Y>0)}] = \Pr(M=1)\Pr(X>0) = E[M]$$
$$E[Y \mid Y>0] = E[M \mid M>0]E[X \mid X>0] = E[X].$$

The first equality cannot hold for the zero-inflated distribution because $\Pr(X>0) \neq 1$. The second equality holds for both zero-inflated and hurdle models because $E[M \mid M>0]$ is identically 1, because $M \sim Bernoulli$. However, $E[X \mid X>0] = E[X]$ is correct only for a hurdle distribution and cannot hold for zero-inflated distribution. In this situation the parameters of the first and second processes of the hurdle distribution can be consistently estimated with the condition $\boldsymbol{G} = (\boldsymbol{G_1}, \boldsymbol{G_2})$, by the generalized method of moment (GMM):

$$E\left[I_{(Y>0)} - E[M]\right] = \boldsymbol{G_1} = \frac{1}{n}\sum g_1(\boldsymbol{a}) = 0 \qquad \text{(A.1)}$$

$$E\left[Y - E[X] \middle| Y>0\right] = \boldsymbol{G_2} = \frac{1}{n}\sum g_2(\boldsymbol{\beta}) = 0. \qquad \text{(A.2)}$$

where $g_1(\boldsymbol{a})$ is a ($p \times 1$) vector and $g_2(\boldsymbol{\beta})$ a $(k \times 1)$ vector. Therefore, having evaluated the parameters of the hurdle distribution by the last two equations using the generalized method of moments, the test can be done by checking the following equality:

$$E\left[Y - E[M]E[X]\right] = \boldsymbol{D} = \frac{1}{n}\sum d(\boldsymbol{a}, \boldsymbol{\beta}) = 0$$

Using $\theta = (\boldsymbol{a}, \boldsymbol{\beta})$, under the null hypothesis that the model is correctly specified and that the separation hypothesis holds, explicit equations of this test are, following the notations of Prieger (2003):

$$T_{CM} = \lim_{n\to\infty} n\boldsymbol{D}'\Sigma_0^{-1}\boldsymbol{D}$$

with:

$$\Sigma_0 = V_{dd'} + D_0 J_0^{-1} V_{gd'} + V_{dg'} J_0^{-1} D_0 + D_0 J_0^{-1} V_{dd'} J_0^{-1} D_0'$$

$$J_0 = -\lim_{n\to\infty} \frac{1}{n}\frac{\delta \boldsymbol{G}}{\delta\theta} = -\lim_{n\to\infty}\frac{1}{n}\begin{bmatrix} \frac{\delta \boldsymbol{G_1}}{\delta a} & 0 \\ 0 & \frac{\delta \boldsymbol{G_2}}{\delta \beta}\end{bmatrix}$$

$$D_0 = \lim_{n\to\infty}\frac{1}{n}\frac{\delta \boldsymbol{D}}{\delta\theta}$$

$$V_{dd'} = \boldsymbol{D}\times\boldsymbol{D}'$$
$$V_{gd'} = \boldsymbol{G}\times\boldsymbol{D}'$$
$$V_{dg'} = \boldsymbol{D}\times\boldsymbol{G}'$$
$$V_{gg'} = \boldsymbol{G}\times\boldsymbol{G}'$$

The conditional moment test T_{CM} is asymptotically $\chi^2(r)$ and the condition is rejected at significance level δ when $T_{CM} > \chi^2(r;\delta)$, where r is the number of tested restrictions in the model. Rejection of the test indicates model misspecification, where the parameters of the distributions are not correctly estimated by equations (A.1) and (A.2). In this case, the null hypothesis is rejected, revealing some kind of misspecification, although not necessarily the *separation hypothesis*.

References

- Abdel-Aty, M., Keller, J. 2005. Exploring the overall and specific crash severity levels at signalized intersections. Accident Analysis and Prevention, 37, 417-425.
- Ayuso, M., Guillén, M., Alcañiz, M. 2010. The impact of traffic violations on the estimated cost of traffic accidents with victims. Accident Analysis and Prevention, 42, 709-717.
- Ayuso, M., Santolino, M. 2007. Predicting automobile claims bodily injury severity with sequential ordered logit models. Insurance: Mathematics and Economics, 4, 71-83.
- Baughman, A.L. 2007. Mixture model framework facilities understanding of zero-inflated and hurdle models for count data. Journal of Biopharmaceutical Statistics, 17, 943-946.
- Boucher, J.-P., Denuit, M., Guillén, M. 2007. Risk classification for claim counts: A comparative analysis of various zero-inflated mixed Poisson and hurdle models. North American Actuarial Journal, 11(4), 110-131.
- Boucher, J.-P., Denuit, M., Guillén, M. 2009. Number of accidents or number of claims? an approach with zero-inflated Poisson models for panel data. Journal of Risk and Insurance, 76, 821-846.
- Boucher, J.-P., Guillén, M. 2008.A survey on models for panel count data with applications to insurance. Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales, 103,2, 277-295.
- Cameron, A.C., Trivedi, P.K. 1998. Regression analysis of count data. New York: Cambridge University Press.
- Cragg, J. 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica, 39(5), 829--844.
- Delen, D., Sharda, R., Bessonov, M. 2006. Identifyng significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis and Prevention, 38, 434-444.
- Dionne, G., Vanasse, C. 1989. A generalization of automobile insurance rating models:

the Negative Binomial distribution with Regression Component. ASTIN Bulletin, 19, 199-212.

- EC 2003. Draft report with recommendations to the Commission on a European disability rating scale. Committee on Legal Affairs and the Internal Market 2003/2130(INI).
- Eluru, N., Bhat, C., Hensher, D. 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis and Prevention, 40, 1033-1054.
- Eluru and Bhat, 2007. A joint econometric analysis of seat belt use and crash related injury severity. Accident Analysis and Prevention, 39, 1037-1049.
- Evans, L. 2001. Female compared with male fatality risk from similar physical impacts. The Journal of Trauma: Injury, Infection, and Critical Care, 50, 281-288.
- Greenwood, M., Yule, G. 1920. An inquiry into the nature of frequency distributions representative of multiple happenings with particular references to the occurence of multiple attacks of disease or repeated accidents. Journal of the Royal Statistical Society, 83, 255-279.
- Grootendorst, P. 1995. A comparison of alternative models of prescription drug utilization. Health Economics. 4,183-198.
- Gurmu, S. 1998. Generalized hurdle count data regression models. Economics Letters, 58, 263-268.
- Huang, H., Chin, H.-C., Haque, M.-M. 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: A Bayesian hierarchical analysis. Accident Analysis and Prevention, 40, 45-54.
- Holubowycz, O.T. 1995. Age, sex, and blood alcohol concentration of killed and injured pedestrians. Accident Analysis and Prevention, 27, 417-422.
- Johnson, N., Kotz, S., Balakrishnan, N. 1995. Continuous univariate distributions. New York: Wiley, 2nd ed.
- Jung, S., Qin, X., Noyce, D. A. 2010. Rainfall effect on single-vehicle crash severities using polychotomous response models. Accident Analysis and Prevention, 42, 213-224.
- Khattak, A.J., Rocha, M. 2003. Are SUVs `supremely unsafe vehicles'? Analysis of

rollovers and injuries with sport utility vehicles. Transportation Research Record, 1840, 167-177.

- Kockelman, K.M., Kweon, Y.-J. 2002. Driver injury severity: an application of ordered probit models. Accident Analysis and Prevention, 34, 313-321.
- Lambert, D. 1992. Zero-inflated Poisson regression with an application to defects in manufacturing. Technometrics, 34, 1-14.
- Lawless, J. 1987. Negative Binomial and mixed Poisson regression. The Canadian Journal of Statistics, 15, 209-225.
- Lee, C., Abdel-Aty, M. 2008. Presence of passengers: Does it increase or reduce driver's crash potential? Accident Analysis and Prevention, 40, 1703-1712.
- Lord, D., Washington, S., Ivan, J. 2005. Poisson, Poisson-Gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. Accident Analysis and Prevention, 37, 35-46.
- Lord, D., Washington, S., Ivan, J. 2007. Further notes on the application of zero-inflated models in highway safety. Accident Analysis and Prevention, 39, 53-57.
- Malyshkina, N., Mannering, F. 2009. Markov switching multinomial logit model: An application to accident-injury severities. Accident Analysis and Prevention, 41, 829-838.
- Milton, J., Shankar, V., Mannering, F. 2008. Highway accident severities and the mixed logit model: An explanatory empirical analysis. Accident Analysis and Prevention, 40, 206-266.
- Mullahy, J. 1986. Specification and testing in some modified count data models. Journal of Econometrics, 33, 341-365.
- Newey, W. 1985. Generalized method of moment specification testing. Journal of Econometrics, 29, 229-256.
- O'Donnell, C.J., Connor, D.H. 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis and Prevention, 28, 739-753.
- Pohlmeier, W., Ulrich, V. 1995. An econometric model of the two-part decision making process in the demand for health care. Journal of Human Resources, 30, 339-361.

- Prieger, J.E. 2003, Bootstrapping the conditional moment tests for parametric duration models. Applied Economics Letters, 10, 597-600.
- Rueda-Domingoa, T., Lardelli-Claret, P., Luna-del-Castillo, J., Jiménez-Moleón, J.J., García-Martín, M., Bueno-Cavanillas, A., 2004. The influence of passengers on the risk of the driver causing a car collision in Spain: analysis of collisions from 1990 to 1999. Accident Analysis and Prevention, 36, 481-489.
- Santos-Silva, J., Windmeijer, F. 2001. Two-part multiple spell models for health care demand. Journal of Econometrics, 104, 67-89.
- Savolainen, P., Mannering, F. 2007. Probabilistic models of motorcyclists' injury severities in single-and multi-vehicle crashes. Accident Analysis and Prevention, 39, 955-963.
- Srinivasan, K.K. 2002. Injury severity analysis with variable and correlated thresholds: ordered mixed logit formulation. Transportation Research Record, 1784, 132-142.
- Tauchen, G. 1985. Diagnostic testing and evaluation of maximum likelihood models. Journal of Econometrics, 30, 415-443.
- Ulfarsson, G.F., Mannering, F.L. 2004. Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents. Accident Analysis and Prevention, 36, 135-147.
- Ullah, S., Finch, C., Day, L. 2009. Statistical modelling for falls count data. Accident Analysis and Prevention, doi:10.1016/j.aap.2009.08.018
- Vuong, Q. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica, 57, 307-333.
- Wang, X., Kockelman, K.M. 2005. Use of heteroscedastic ordered logit model to study severity of occupant injury: Distinguishing effects of vehicle weight and type. Transportation Research Record 1908, 195-204.
- Winkelmann, R. 2003a. Health care reform and the number of doctor visits - an econometric analysis. Journal of Applied Econometrics, 19, 455-472.

- Winkelmann, R. 2003b. Econometric analysis of count data. Springer-Verlag, Berlin, 4th ed.
- Yamamoto, T., Hashiji, J., Shankar, V. 2008. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. Accident Analysis and Prevention, 40, 1320-1329.

Llista Document de Treball
List Working Paper

WP 2010/05 "Discrete distributions when modeling the disability severity score of motor victims"Jean-Philippe Boucher and Miguel Santolino

WP 2010/04 "Does privatization spur regulation? Evidence from the regulatory reform of European airports . Bel, G. and Fageda, X."

WP 2010/03 "High-Speed Rail: Lessons for Policy Makers from Experiences Abroad". Albalate, D ; and Bel, G."

WP 2010/02 "Speed limit laws in America: Economics, politics and geography". Albalate, D ; and Bel, G."

WP 2010/01 "Research Networks and Inventors' Mobility as Drivers of Innovation: Evidence from Europe" Miguélez, E. ; Moreno, R. "

WP 2009/26 "Social Preferences and Transport Policy: The case of US speed limits" Albalate, D.

WP 2009/25 "Human Capital Spillovers Productivity and Regional Convergence in Spain" , Ramos, R ; Artis, M. ; Suriñach, J.

WP 2009/24 "Human Capital and Regional Wage Gaps" ,López-Bazo,E. Motellón E.

WP 2009/23 "Is Private Production of Public Services Cheaper than Public Production? A meta-regression analysis of solid waste and water services" Bel, G.; Fageda, X.; Warner. M.E.

WP 2009/22 "Institutional Determinants of Military Spending" Bel, G., Elias-Moreno, F.

WP 2009/21 "Fiscal Regime Shifts in Portugal" Afonso, A., Claeys, P., Sousa, R.M.

WP 2009/20 "Health care utilization among immigrants and native-born populations in 11 European countries. Results from the Survey of Health, Ageing and Retirement in Europe" Solé-Auró, A., Guillén, M., Crimmins, E.M.

WP 2009/19 "La efectividad de las políticas activas de mercado de trabajo para luchar contra el paro. La experiencia de Cataluña" Ramos, R., Suriñach, J., Artís, M.

WP 2009/18 "Is the Wage Curve Formal or Informal? Evidence for Colombia" Ramos, R., Duque, J.C., Suriñach, J.

WP 2009/17 "General Equilibrium Long-Run Determinants for Spanish FDI: A Spatial Panel Data Approach" Martínez-Martín, J.

WP 2009/16 "Scientists on the move: tracing scientists' mobility and its spatial distribution" Miguélez, E.; Moreno, R.; Suriñach, J.

WP 2009/15 "The First Privatization Policy in a Democracy: Selling State-Owned Enterprises in 1948-1950 Puerto Rico" Bel, G.

WP 2009/14 "Appropriate IPRs, Human Capital Composition and Economic Growth" Manca, F.

WP 2009/13 "Human Capital Composition and Economic Growth at a Regional Level" Manca, F.

WP 2009/12 "Technology Catching-up and the Role of Institutions" Manca, F.

WP 2009/11 "A missing spatial link in institutional quality" Claeys, P.; Manca, F.

WP 2009/10 "Tourism and Exports as a means of Growth" Cortés-Jiménez, I.; Pulina, M.; Riera i Prunera, C.; Artís, M.

WP 2009/09 "Evidence on the role of ownership structure on firms' innovative performance" Ortega-Argilés, R.; Moreno, R.

WP 2009/08 "¿Por qué se privatizan servicios en los municipios (pequeños)? Evidencia empírica sobre residuos sólidos y agua" Bel, G.; Fageda, X.; Mur, M.

WP 2009/07 "Empirical analysis of solid management waste costs: Some evidence from Galicia, Spain"Bel, G.; Fageda, X.

WP 2009/06 "Intercontinental fligths from European Airports: Towards hub concentration or not?" Bel, G.; Fageda, X.

WP 2009/05 "Factors explaining urban transport systems in large European cities: A cross-sectional approach" Albalate, D.; Bel, G.

WP 2009/04 "Regional economic growth and human capital: the role of overeducation" Ramos, R.; Suriñach, J.; Artís, M.

WP 2009/03 "Regional heterogeneity in wage distributions. Evidence from Spain" Motellón, E.; López-Bazo, E.; El-Attar, M.

WP 2009/02 "Modelling the disability severity score in motor insurance claims: an application to the Spanish case" Santolino, M.; Boucher, J.P.

WP 2009/01 "Quality in work and aggregate productivity" Royuela, V.; Suriñach, J.

WP 2008/16 "Intermunicipal cooperation and privatization of solid waste services among small municipalities in Spain" Bel, G.; Mur, M.

WP 2008/15 "Similar problems, different solutions: Comparing refuse collection in the Netherlands and Spain" Bel, G.; Dijkgraaf, E.; Fageda, X.; Gradus, R.

WP 2008/14 "Determinants of the decision to appeal against motor bodily injury settlements awarded by Spanish trial courts" Santolino, M

WP 2008/13 "Does social capital reinforce technological inputs in the creation of knowledge? Evidence from the Spanish regions" Miguélez, E.; Moreno, R.; Artís, M.

WP 2008/12 "Testing the FTPL across government tiers" Claeys, P.; Ramos, R.; Suriñach, J.

WP 2008/11 "Internet Banking in Europe: a comparative analysis" Arnaboldi, F.; Claeys, P.

WP 2008/10 "Fiscal policy and interest rates: the role of financial and economic integration" Claeys, P.; Moreno, R.; Suriñach, J.

WP 2008/09 "Health of Immigrants in European countries" Solé-Auró, A.; M.Crimmins, E.

WP 2008/08 "The Role of Firm Size in Training Provision Decisions: evidence from Spain" Castany, L.

WP 2008/07 "Forecasting the maximum compensation offer in the automobile BI claims negotiation process" Ayuso, M.; Santolino, M.

WP 2008/06 "Prediction of individual automobile RBNS claim reserves in the context of Solvency II" Ayuso, M.; Santolino, M.

WP 2008/05 "Panel Data Stochastic Convergence Analysis of the Mexican Regions" Carrion-i-Silvestre, J.L.; German-Soto, V.

WP 2008/04 "Local privatization, intermunicipal cooperation, transaction costs and political interests: Evidence from Spain" Bel, G.; Fageda, X.

WP 2008/03 "Choosing hybrid organizations for local services delivery: An empirical analysis of partial privatization" Bel, G.; Fageda, X.

WP 2008/02 "Motorways, tolls and road safety. Evidence from European Panel Data" Albalate, D.; Bel, G.

WP 2008/01 "Shaping urban traffic patterns through congestion charging: What factors drive success or failure?" Albalate, D.; Bel, G.

WP 2007/19 "La distribución regional de la temporalidad en España. Análisis de sus determinantes" Motellón, E.

WP 2007/18 "Regional returns to physical capital: are they conditioned by educational attainment?" López-Bazo, E.; Moreno, R.

WP 2007/17 "Does human capital stimulate investment in physical capital? evidence from a cost system framework" López-Bazo, E.; Moreno, R.

WP 2007/16 "Do innovation and human capital explain the productivity gap between small and large firms?" Castany, L.; López-Bazo, E.; Moreno, R.

WP 2007/15 "Estimating the effects of fiscal policy under the budget constraint" Claeys, P.

WP 2007/14 "Fiscal sustainability across government tiers: an assessment of soft budget constraints" Claeys, P.; Ramos, R.; Suriñach, J.

WP 2007/13 "The institutional vs. the academic definition of the quality of work life. What is the focus of the European Commission?" Royuela, V.; López-Tamayo, J.; Suriñach, J.

WP 2007/12 "Cambios en la distribución salarial en españa, 1995-2002. Efectos a través del tipo de contrato" Motellón, E.; López-Bazo, E.; El-Attar, M.

WP 2007/11 "EU-15 sovereign governments' cost of borrowing after seven years of monetary union" Gómez-Puig, M..

WP 2007/10 "Another Look at the Null of Stationary Real Exchange Rates: Panel Data with Structural Breaks and Cross-section Dependence" Syed A. Basher; Carrion-i-Silvestre, J.L.

WP 2007/09 "Multicointegration, polynomial cointegration and I(2) cointegration with structural breaks. An application to the sustainability of the US external deficit" Berenguer-Rico, V.; Carrion-i-Silvestre, J.L.

WP 2007/08 "Has concentration evolved similarly in manufacturing and services? A sensitivity analysis" Ruiz-Valenzuela, J.; Moreno-Serrano, R.; Vaya-Valcarce, E.

WP 2007/07 "Defining housing market areas using commuting and migration algorithms. Catalonia (Spain) as an applied case study" Royuela, C.; Vargas, M.

WP 2007/06 "Regulating Concessions of Toll Motorways, An Empirical Study on Fixed vs. Variable Term Contracts" Albalate, D.; Bel, G.

WP 2007/05 "Decomposing differences in total factor productivity across firm size" Castany, L.; Lopez-Bazo, E.; Moreno, R.

WP 2007/04 "Privatization and Regulation of Toll Motorways in Europe" Albalate, D.; Bel, G.; Fageda, X.

WP 2007/03 "Is the influence of quality of life on urban growth non-stationary in space? A case study of Barcelona" Royuela, V.; Moreno, R.; Vayá, E.

WP 2007/02 "Sustainability of EU fiscal policies. A panel test" Claeys, P.

WP 2007/01 "Research networks and scientific production in Economics: The recent spanish experience" Duque, J.C.; Ramos, R.; Royuela, V.

WP 2006/10 "Term structure of interest rate. European financial integration" Fontanals-Albiol, H.; Ruiz-Dotras, E.; Bolancé-Losilla, C.

WP 2006/09 "Patrones de publicación internacional (ssci) de los autores afiliados a universidades españolas, en el ámbito económico-empresarial (1994-2004)" Suriñach, J.; Duque, J.C.; Royuela, V.

WP 2006/08 "Supervised regionalization methods: A survey" Duque, J.C.; Ramos, R.; Suriñach, J.

WP 2006/07 "Against the mainstream: nazi privatization in 1930s germany" Bel, G.

WP 2006/06 "Economía Urbana y Calidad de Vida. Una revisión del estado del conocimiento en España" Royuela, V.; Lambiri, D.; Biagi, B.

WP 2006/05 "Calculation of the variance in surveys of the economic climate" Alcañiz, M.; Costa, A.; Guillén, M.; Luna, C.; Rovira, C.

WP 2006/04 "Time-varying effects when analysing customer lifetime duration: application to the insurance market" Guillen, M.; Nielsen, J.P.; Scheike, T.; Perez-Marin, A.M.

WP 2006/03 "Lowering blood alcohol content levels to save lives the european experience" Albalate, D.

WP 2006/02 "An analysis of the determinants in economics and business publications by spanish universities between 1994 and 2004" Ramos, R.; Royuela, V.; Suriñach, J.

WP 2006/01 "Job losses, outsourcing and relocation: empirical evidence using microdata" Artís, M.; Ramos, R.; Suriñach, J.

Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

WEBSITE: www.ub.edu/irea/ • **CONTACT:** irea@ub.edu