UNIVERSITAT ROVIRA I VIRGILI

WORKING PAPERS

Col·lecció "DOCUMENTS DE TREBALL DEL DEPARTAMENT D'ECONOMIA"

"Locations and Relocations: Modelling, Determinants, and Interrelations"

Josep Maria Arauzo Carod
Miguel C. Manjón Antolín

Document de treball nº -6- 2007

DEPARTAMENT D'ECONOMIA
Facultat de Ciències Econòmiques i Empresarials

Edita:

Departament d'Economia
http://www.fcee.urv.es/departaments/economia/public_html/index.html
Universitat Rovira i Virgili
Facultat de Ciències Econòmiques i Empresarials
Avgda. de la Universitat, 1
432004 Reus
Tel. +34 977 759 811
Fax +34 977 300 661

Dirigir comentaris al Departament d'Economia.

Dipòsit Legal: T - 1886 - 2007
ISSN 1988 - 0812

DEPARTAMENT D'ECONOMIA
Facultat de Ciències Econòmiques i Empresarials

Locations and Relocations:

Modelling, Determinants, and Interrelations*

Miguel C. Manjón-Antolín**

(QURE - Department of Economics, Rovira i Virgili University)

and

Josep-Maria Arauzo-Carod

(QURE - Department of Economics, Rovira i Virgili University)

---

* We are grateful to J. Trullén and R. Boix for kindly providing their municipalities' database and to D. Padro for his help in dealing with the Fortram program of Professor Krishnamoorthy. Thanks are also due to W. Boente, P. McCann, T. Mora, I. Mariotti, L. Moreno, I. Peña and seminar participants at the University of Barcelona (CAEPS), the Max Planck Institute of Economics (Jena), the 46th ERSA Conference (Volos), the XXII "Jornadas de Economía Industrial" (IESE, Barcelona), the University of Valencia (Department of Applied Economics II) and the "XXXII Simposio de Análisis Económico" (Granada) for their useful comments on early drafts of this paper. We also acknowledge the institutional support of the CIDEM. This paper is part of a research project financially supported by grants SEJ2007-64605/ECON and SEJ2007-65086/ECON as well as by the "Xarxa de Referència d'R+D+I en Economia i Polítiques Públiques" of the Catalan Government. Any errors are of course our own.

** Corresponding author (miguel.manjon@urv.net). Address for correspondence: Faculty of Economics and Business, Av. Universitat 1, Reus-43204, Spain.

Locations and Relocations:
Modelling, Determinants, and Interrelations

Abstract

Empirical studies on industrial location do not typically distinguish between new and relocated establishments. This paper addresses this shortcoming using data on the frequency of these events in municipalities of the same economic-administrative region. This enables us to test not only for differences in their determinants but also for interrelations between start-ups and relocations. Estimates from count regression models for cross-section and panel data show that, although partial effects differ, common patterns arise in "institutional" and "neoclassical" explanatory factors. Also, start-ups and relocations are positive but asymmetrically related.

JEL classification: C25, R30, R10
Keywords: cities, count data models, industrial location

1. Introduction

There is extensive empirical literature on the determinants of industrial location (see e.g. Arauzo and Manjón 2007). However, most of this research implicitly assumes that start-ups are all similar. This assumption contrasts with evidence that shows that start-ups significantly differ in features such as e.g. size (Carlton 1979), technology (Frenkel 2001) and geographical origin (Figueiredo et al. 2002). More importantly, these studies conclude that these differences critically affect the way location decisions are taken.

In this paper we examine yet another feature that has comparatively received less attention. Namely, the fact that some start-ups are actually relocations, i.e. businesses established in the past that at some point decide to abandon their current location and move to another one. The distinction is relevant because location decisions are taken on the grounds of incomplete information about the sites and so previous experiences can make a difference. Thus, the opening of new concerns and the relocation of existing concerns are different location processes that should be studied separately (Pellenbarg et al. 2002a, 2002b; Mariotti 2005; Lee 2006).

Accordingly, we analyse the determinants of industrial location distinguishing between strictly-new and relocated concerns. Also, we explore the possibility that both processes, locations and relocations, are interrelated. From an economic policy point of view, these are issues worth inspecting given the increasingly large amount of public funds invested in public incentive programs aiming to attract new businesses (Lee 2004). We use the establishment and the municipality as units of analysis.[1]

This paper is related to the Foreign Direct Investments (FDI) literature to the extent that we share an interest in the relocation of production (see e.g. Blonigen 2005). In contrast with the typical FDI paper,

[1] More specifically, we use data from the "Registry of Industrial Establishments of Catalonia" on plants (i.e. manufacturing establishments) located in the 946 Catalan municipalities in the period 2001-2004. Catalonia is a Spanish region (NUTS-II) in the northeast of Spain whose capital is Barcelona. It has an area of 31,895 km^2, a population of about 7 million people (around 15% of the population of Spain) and its GDP is approximately 20% of Spain's GDP. The reason for choosing Catalonia (instead of e.g. any other Spanish region) to carry out this study was

however, relocations in our dataset typically occur within the same economic-administrative region.[2] This paper also differs from previous studies on the firm/plant relocation decision within the same country in that we are mainly concerned with how the characteristics of a municipality affect the rate of occurrence of start-ups and relocations, rather than with how firm/plant (and, in some studies, location) characteristics affect the decision to relocate and/or the choice of a particular municipality from a set of potential (re)locations.[3]

The closest study in this respect is that of Holl (2004a), who analyses the spatial patterns of start-ups and relocations in Portugal between 1986 and 1997. Because of the integer nature of our dependent variable, the number of new and relocated establishments in a municipality, we both resort to count data models. However, there are major differences with our study. First, she studies a larger geographical area that covers many potentially heterogeneous regions (although some of the covariates she uses may account for this). Second, she uses a set of explanatory variables that, notwithstanding differences in sources and definitions, is largely nested in our specification (e.g. she does not include proxies for "institutional" factors). Third, she does not explore alternative assumptions on the data generation process as we do. Fourth, she does not address the interrelation between start-ups and relocations.[4]

We find that the determinants of start-ups and relocations are practically the same. These include "neoclassical" and "institutional" factors such as e.g. (dis)urbanisation and location economies (neoclassical factors) as well as dummies for the administrative and spatial organisation of the territory (institutional factors). However, the partial derivatives of the conditional expectation of start-ups and

---

the richness of the municipality data-base kindly provided by Trullén and Boix (2005), which we complemented with data from the Catalan Institute of Statistics to construct our vector of explanatory variables.

2 This is just an assumption based on the observed pattern that most relocations follow (see e.g. Stam 2007). Unfortunately, we cannot test it because we do not have data on either the nationality of the firm or the municipality from which relocations come from. In any case, this is not a major concern given the aim of the paper.

3 Cooke (1983), van Dijk and Pellenbarg (2000), Brouwer et al. (2004) and Lee (2004) are examples of studies that analyse the decision to relocate; Baudewyns et al. (2000) and Strauss-Kahn and Vives (2005) are examples of studies that analyse the choice of a relocating site.

4 Moses and Williamson (1967) and Erickson and Wasylenko (1980) are related studies that analyse intra-metropolitan relocations using an analogous dependent variable. However, the former estimate their model by OLS whereas the latter transform the dependent variable into a proportion of the total number of relocating firms and estimate the resulting logistic specification by weighted least squares.

relocations with respect to these determinants, i.e. the partial or marginal effects, differ. Hence, our results do not fully concur with Holl's main "finding (…) that plant start-ups and relocations are not attracted by the same set of location characteristics" (p. 665). Also, we provide evidence of a positive relation between the rates of occurrence of these events. That is, the likelihood that new and previously established concerns (re)locate in a particular municipality increases with the presence of relocations and start-ups, respectively, in that municipality.

The rest of the paper is organised as follows. Section 2 discusses the econometric methodology. Section 3 briefly reviews the literature on the determinants of industrial (re)location. Given that we estimate reduced-form models this review provides useful guidelines for the selection of the explanatory variables. Section 4 contains the empirical results. We describe the data and present inferences from count regression models. Section 5 is the conclusion.

2. Econometric modelling

Discrete choice models (DCM) and count data models (CDM) are the basic econometric tools in empirical studies of industrial location (Arauzo and Manjón 2007). One reason for this is that these models are consistent with a profit maximization framework in which firms choose the optimal location subject to standard constraints —see e.g. Becker and Henderson (2000) and Guimarães et al. (2004). However, DCM critically differ from CDM in the type of data they require and the type of inferences they provide. On the one hand, the unit of analysis in DCM is the firm or the establishment and the main concern is how certain characteristics of this unit (size, sector, etc.) and/or the chosen territory (population, infrastructures, etc.) affect location decisions. On the other hand, the unit of analysis in CDM is geographical (municipality, county, province, region, etc.) and the factors that may affect location decisions refer accordingly to the territory.

In light of these features one could argue that DCM have an advantage over CDM because they may account for both firm and spatial factors. However, there are other issues worth considering when it comes to selecting a model for our location study. One should be aware, for example, that computation of the likelihood function in DCM is cumbersome when the number of alternatives, i.e. sites, is large. Moreover, the set of alternatives in DCM only includes those locations effectively chosen, since the rest do not contribute to the likelihood function. These drawbacks can turn CDM into our preferred specification, especially whenever it is possible to recover the parameter estimates of DCM from the estimates of CDM and/or the sample contains a substantial number of locations where not a single business started operations in the period of analysis (Guimarães et al. 2003). Computational burden is not an issue in CDM and zero observations not only contribute to the likelihood function but provide interesting insights about the data generation process (Mullahy 1997).

CDM seem therefore particularly useful for investigations using highly geographically disaggregated data. We can mention, among others, the urban studies of Holl (2004b) on Portuguese municipalities, Holl (2004c) on Spanish municipalities and Arauzo and Manjón (2004) and Arauzo (2005) on Catalan municipalities. Yet there are studies that apply CDM to larger geographical units, such as e.g. Becker and Henderson (2000), List and McHome (2000) and Guimarães et al. (2004) on US counties and Papke (1991) on US states.[5]

All these studies, however, do not distinguish between strictly-new concerns and relocations. To the best of our knowledge, the only previous study that makes such a distinction and uses CDM is Holl (2004a). She analyses the determinants of Portuguese plant start-ups and relocations using Fixed-Effects estimators for Poisson and Negative Binomial models to control for unobserved municipality-specific heterogeneity. In this paper we essentially follow the same approach. However, we explore alternative specifications to cope with the distinct characteristics of our data. Namely, “excess of zeros” and overdispersion, possible discrepancy between the period of occurrence (exposure) and the period of observation, and dependence between the events of interest. Next we discuss these in detail.[6]

2.1 Excess of zeros and overdispersion

---

[5] CDM are also very popular in the FDI literature —see e.g. Blonigen (2005).

As Figure 1 shows, there are more zeros in the histograms of start-ups and relocations than the Poisson density (with unitary mean) predicts. This largely disqualifies the standard Poisson regression model as a suitable specification for our data. Also, descriptive statistics in Panel A of Table 1 indicate that the conditional variance of our processes exceeds their conditional mean. To see this, one should note that although the conditional variance in Poisson regression models is generally smaller than the non-conditional variance, the conditional expectation should not differ considerably from the sample mean (as long as the model has a constant term). Since "equidispersion", i.e. equality of conditional variance and mean, is one the main assumptions of the Poisson regression model, its rejection further supports the need for less restrictive, more efficient models. One such model is the so-called "mixture model".

[Insert Figure 1 about here]

Mixture models essentially differ from the Poisson regression model in that we introduce a stochastic term ξ in the conditional mean function of the dependent variable, usually in multiplicative form. In standard notation, $E(y|x,\xi) = \exp(x\beta)\xi = \mu\xi$. Within this basic framework, two large classes of CDM arise depending on whether ξ is considered a continuous or discrete variable. (i) In continuous mixtures, ξ has a natural interpretation as an individual random effect that allows for unobserved heterogeneity in the model. That is, ξ aims to account for characteristics of the municipality that are not observed by the researcher (e.g. the "business climate") and/or differences in municipalities beyond those captured by the explanatory variables. Assuming that ξ has a Gamma distribution with unitary mean and constant variance α, for example, leads to one of the most popular continuous mixtures: the Negative Binomial Model (NBM). In particular, here we use the so-called NB2 Model, which assumes a quadratic function for the conditional variance of the dependent variable, $Var(y|x,\xi) = \mu + \alpha\mu^2$. (ii) In finite mixtures, ξ allows for the existence of a discrete number of heterogeneous groups in the population of interest. In the simplest case this amounts to assuming that the population consists of two groups: municipalities in which there are not nor will be new/relocated concerns (e.g. because they are banned by environmental regulations)

---

[6] The statistical foundations of what follows can be found in e.g. Cameron and Trivedi (1998, 2001).

and municipalities in which there might or might not be new/relocated concerns (i.e. in principle there is nothing that prevents this event from happening although it may not happen in certain circumstances, as, for example, when the municipality is too small, too remote, etc.). To construct the finite mixture model this binary-form of heterogeneity is parameterised using, for example, the logistic transformation and the resulting logit model for the probability of zero entrants in the municipality is mapped into a count model that now only accounts for the positive values of the dependent variable. The Zero Inflated Poisson Model (ZIPM) uses the Poisson model to this end, whereas using NBM generates the Zero Inflated Negative Binomial Model (ZINBM).[7]

This is a convenient way to extend the basic Poisson regression model because mixing strategies based on a multiplicative term induces both an excess of zeros and overdispersion in the distribution (Mullahy 1997). It is important to keep in mind, however, the way in which each model brings this result about. NBM controls for heterogeneous municipalities but assumes the same data generation process for zeros and positive outcomes. In contrast, ZIPM distinguishes two regimes in the data generation process. The downside is that ZIPM does not account for overdispersion in the positive set. ZINBM addresses this downside, though it is less parsimonious than either ZIPM or (of course) NBM. The practical problem is that we cannot easily discern which of these mechanisms is ultimately responsible for the observed excess of zeros and overdispersion.[8]

Several tests and information criteria may help us to select the appropriate model for our data (Andrews 1988, Sin and White 1996, Cameron and Trivedi 1998). To this end we report the value of the log-

---

[7] We use "zero-inflated" rather than, for example, "hurdle" or "two-parts" models (other mixture models are left for future research) because in the latter the finite mixture arises from the combination of a process that generates the zeros with another that generates strictly positive outcomes. This amounts to assuming that the population consists of two groups: municipalities in which there are new/relocated concerns and municipalities in which there are not. Clearly, this is a too restrictive framework for our study. In particular, we use the population of the municipality as the main determinant of the probability of zero start-ups and relocations. In all the specifications used in this study this variable was negative and statistically significant.

[8] This identification problem may be more complex, for failure of the "independence of events" assumption in the Poisson process also causes overdispersion (Cameron and Trivedi 2001: 337). In the context of industrial location studies this assumption means that the likelihood that an establishment locates in a particular municipality is independent of other establishments being located there. This seems a restrictive assumption, although we should not

likelihood function (denoted by "Log L" in the tables of results), the Akaike Information Criterion ("AIC"), the Likelihood-Ratio test for the joint significance of the model ("LR Joint Test") and the χ^2 goodness-of-fit test ("GoF Test").[9] In addition, we can take advantage of the nested structure of the ZIPM in the ZINBM and perform a LR-type test between these models (reported as "LR Inflated Test" in the tables of results) based on the null hypothesis that the parameter α is zero. Finally, the "Vuong Test" provides us with a non-nested testing procedure that discriminates between Poisson and Negative Binomial models and their respectively inflated specifications, ZIPM and ZINBM.

2.2 Exposure

Another characteristic of the data that deserves attention concerns the definition of the exposure. Ideally, the period of observation should be the same as the period of occurrence of the events. However, this is often not the case. Our data, for example, are annually recorded. Yet this does not necessarily mean that the exposure period is annual. In fact, there is no economic or legal reason why the rate of occurrence of start-ups and relocations should be calculated on a yearly basis rather than over, for example, the available four-year period (in general, over any other period). We attempt to deal with this indeterminacy by considering alternative definitions of the exposure period.

We initially assume that the periods of occurrence and observation are the same. Accordingly, the dependent variables are the number of start-ups and relocations reported over the period 2001 to 2004 and the explanatory variables are calculated as period-means. In Table 2 we report estimates from NBM, ZIPM and ZINBM using these data. Later we assume that the period of occurrence is annual, which is the period used by the statistical sources that provide our explanatory variables (the Catalan Institute of Statistics and Trullén and Boix 2005). This lead us to examine two additional questions.

---

pursue this issue here. Notice, however, that the discussion below on the dependence between start-ups and relocations partially addresses this assumption.

[9] To construct the "Gof Test" we use the computational approach proposed by Andrews (1988: Appendix 5). In particular, results presented below were obtained partitioning the data in four cells. We also tried other values for the number of cells, but the main conclusions remained largely unaltered.

The first is whether the yearly rates of occurrence of start-ups and relocations are independent. To answer this question we calculated the covariance matrix for the year vector of Pearson-residuals from the pooled Poisson regression model. As Hausman et al. (1984) argue, if the assumption of time independence holds, the resulting 4×4 matrix should have small values in the off diagonal elements, whilst cross-section estimates would provide valid inferences. The estimated correlations are indeed practically negligible, with highest values of order 0.3 but typically around 0.02. We thus find evidence supporting the time independence property in our data.[10]

The second is whether the data generation process governing the yearly rates of occurrence of start-ups and relocations is the same every year. To answer this question we compared NBM, ZIPM and ZINBM estimates from different sample years. We found that the value of the coefficients slightly varied across years, but their sign and statistical significance was the same regardless of the sample year. We thus find evidence supporting the hypothesis that the data generation process is indeed the same. Consequently, in Table 3 we just report results from the 2001 cross-section. These data provided the smallest AIC in all the specifications.[11]

Still, it may be interesting to take full advantage of the panel structure of our data set and calculate panel data estimates. As discussed in the previous subsection, NBM and ZINBM enable us to control for unobserved municipality-specific heterogeneity in cross section data. However, they are not Poisson models. Panel data estimators also control for such unobserved heterogeneity, but they can do it maintaining the assumption that the data is Poisson distributed. These estimates may therefore be useful to assess the robustness of the conclusions extracted from negative binomial models using cross-section

---

[10] Notice that if the time independence assumption does not hold, cross-section estimators cannot "distinguish (…) between true time independence versus apparent dependence due to the unobserved heterogeneity of the individual units" (Hausman et al. 1984: 911). In this context panel data estimators are needed to control for the serial correlation induced by the presence of individual effects. This is the (untested) assumption implicitly made by e.g. Holl (2004a, 2004b, 2004c) and List and McHome (2000); see, in contrast, Papke (1991), Becker and Henderson (2000) and Guimarães et al. (2004).

[11] One might argue that these results could be distorted by the fact that some of our explanatory variables refer to the year 2001 (see Table 1). However, Sin and White (1996) demonstrate that the AIC may select the appropriate model even when comparing between overlapping, nested and/or misspecified models.

data. On the other hand, panel data estimators based on the Poisson distribution impose equidispersion. As in the cross-section case, Negative Binomial specifications can cope with this.

In Table 4 we accordingly report results from Poisson and Negative Binomial fixed- and random-effects estimators (hereafter FE and RE, respectively). The issue, though, is not whether the latent effects are considered fixed or random but what is their stochastic relation with the covariates, for it is this relation that determines the statistical properties of the estimators. The RE estimator is not consistent if the covariates are correlated with the effects, whereas the FE estimator is consistent regardless of the correlation between covariates and effects. Moreover, zero correlation between covariates and latent effects renders the RE estimator efficient. These properties match the general conditions of the "Hausman Test" (reported in Table 4), which may consequently help us to choose between FE and RE. Statistical properties notwithstanding, we should also take into account that time-invariant explanatory variables are not identified in the conditional (on a sufficient statistic) maximum likelihood framework of the FE estimator. Since a good deal of our explanatory variables is of this kind, we have decided, regardless of the results of the Hausman test, to report both FE and RE estimates in Table 4. However, the latter should be interpreted with care in those cases where the null hypothesis of independence between covariates and individual effects is rejected.

2.3 Dependence

Our last concern with the data relates to the dependence between the events of interest. We have argued that the openings of new and relocated establishments are different location processes, but so far we have implicitly assumed that they are totally independent. Here we discuss whether this is a reasonable assumption.

A positive relation may arise if, for example, start-ups/relocations interpret the increase/decrease of relocations/start-ups in a municipality as a sign that this municipality is a better/worse location. The relationship would also be positive if, for example, start-ups and relocations in a municipality are up(down)stream producers of a vertically organised industry (that is, when start-ups are created to provide services and products to recently relocated agents, and when relocated businesses are main suppliers to start-ups and decide to move closer to their consumers to save costs, improve contacts, etc.). However, a negative relation may also be possible since, for example, start-ups and relocations may see each other as powerful competitors that may push them out of the local market. Under these circumstances an increase of relocations/start-ups in a municipality may lead potential start-ups/relocations to sensibly opt for installing their establishments in other locations (or simply postpone the decision). But the relation would also be negative if, for example, start-ups saw a decrease in relocations in a municipality as an opportunity for making businesses, or vice versa.

These examples obviously do not constitute solid proof of the existence of dependence, but they do suggest that “signalling” and/or “vertical integration” effects (positive dependence) as well as “rejection” and/or “opportunity” effects (negative dependence) may exist between start-ups and relocations. It is therefore a matter of empirical research to verify this tenet. The problem is that since these effects are not mutually exclusive, i.e. they may occur simultaneously within a municipality, we cannot fully identify them using CDM as the econometric specification and the municipality as the unit of analysis. In this context the best we can do is to test whether on average positive effects dominate negative effects or vice versa. More specifically, we can test whether the likelihood that businesses located in a particular municipality move to another one is, ceteris paribus, positively/negatively affected by the observation that such a location has been chosen for starting up new businesses and whether the likelihood that new business start-up activities in a particular municipality is ceteris paribus, positively/negatively affected by the observation that this municipality has been chosen for relocating existing concerns.

A simple way to implement this idea in our specifications is to include as an additional determinant of the number of start-ups (relocations) the number of relocations (start-ups) and interpret the sign and significance of the associated coefficients as evidence of positive/negative dependence in a typical municipality. At first sight such a specification looks like a system of equations, to be estimated using either a full (maximum likelihood) or a limited (e.g. GMM) information approach. However, the previous discussion suggests that the interrelation between start-ups and relocations is more of a sequential than simultaneous nature. Moreover, "[f]ully specified simultaneous equations models for counts have not been yet developed" (Cameron and Trivedi, 2001: 342). Accordingly, rather than specifying the joint distribution of the variables of interest, we shall analyse dependence by estimating the (mean of the) conditional distribution using the previously described methods (NBM, ZIPM and ZINBM).[12]

In particular, we will proceed in an analogous way as when we assumed that the periods of occurrence and observation are the same. Firstly, we will restrict the sample to the 2002 to 2004 period and include the number of start-ups and relocations in 2001 among the (exogenous) determinants of relocations and start-ups, respectively. These results are reported in Table 5. Secondly, we will restrict the sample to the 2002 cross-section and include the number of start-ups and relocations in 2001 among the (exogenous) determinants of relocations and start-ups, respectively. These results are reported in Table 6.[13]

3. The determinants of industrial (re)location

Having discussed the econometric modelling we turn now to the selection of the explanatory variables. To this end, this section briefly reviews the theoretical and empirical literature on the determinants of industrial (re)location. This will provide guidelines for the construction of our specifications as well as reveal the differences and similarities between our investigation and previous related studies.

3.1 Theories

The location of economic activity has been analysed from a wide range of theoretical perspectives. None of them, however, has dedicated much effort to investigate the idiosyncrasies of relocations. As Brouwer et al. (2004: 336) point out, "[r]elocation theories are hardly applied and are often treated as a special case of location theories". In any case, a thorough discussion of each theory is clearly beyond the scope of this

---

[12] It is interesting to note that, assuming that the data generation process is Poisson distributed and plants' choices are only determined by sites characteristics following a conditional logit model, in this setting it is also possible to identify dependence at the establishment level (see Case 1 in Guimarães et al. 2003).

[13] We also run analogous regression using alternative periods and lags. However, these specifications were the best in terms of likelihood and penalized likelihood criteria.

paper. For our purposes it suffices to outline their principal tenets. In particular, we follow Hayter (1997) in distinguishing three main approaches in location theory: neoclassical, behavioural and institutional.

The decision setting in the neoclassical theory involves rational agents choosing optimally a site among a set of finite alternatives. Hence, in this framework the main determinants of industrial location are those affecting the expected benefits derived from the decision to locate in a particular site. These include, for example, transportation and labour costs, external economies and market size. Most of the studies discussed in the previous section are largely based on these (constrained) profit-maximisation or cost-minimising strategies —see e.g. Frenkel (2001) and Guimarães et al. (2003, 2004).

As for the behavioural theory, it stems essentially from the same decision setting as that of the neoclassical theory. However, it calls into question the assumptions of rationality and perfect information. Instead, agents have limited knowledge and take their location decisions in a world of uncertainty. Unlike the neoclassical approach, which places great emphasis on "external" factors, the behavioural approach stresses the importance of "internal" (size, age, etc.) and "entrepreneurial" (previous experience, residence, etc.) factors in the location decision. Supportive evidence shows, for example, that large firms tend to consider wider sets of alternatives than small firms (Arauzo and Manjón 2004) and that entrepreneurs are more likely to choose locations near her/his residence (Figueiredo et al. 2002).

Lastly, the institutional theory disagrees with the notion held by neoclassical and behavioural theories that firms are isolated agents. In fact, this theory notes that it is quite the opposite: firms operate within a network of clients, suppliers, competitors, trade unions, regional systems, governments, etc. The environment thus matters and should consequently be taken into account when modelling location decisions. Accordingly, institutional theory advocates paying more attention to issues such as e.g. wages, unionisation and regulations. Carlton (1983), Papke (1991) and List and McHome (2000), for example, provide empirical evidence on some of these issues.

All in all, these theories provide a solid analytical framework for a variety of research questions in industrial location. But are they that useful for studying relocations? Pellenbarg et al. (2002a: 11) contend that location theories do "provide the theoretical background for studies of firm relocation". This is also implicitly maintained by van Dijk and Pellenbarg (2000), Brouwer et al. (2004) and Holl (2004a), who all refer to neoclassical, behavioural and institutional theories to motivate their empirical investigations.

However, a caveat is in order: the forces driving location and relocation processes differ. This is not always apparent because location theories tend to overemphasise (minimise) the importance of "pull" ("push") factors in relocation decisions. However, relocations differ from strictly new locations in that they are the outcome of a sequence of decisions taken over the history of the firm or the establishment. That is, relocation decisions are taken conditionally upon previous location decisions (Pellenbarg et al. 2002b). It is therefore plausible to conclude that the information used to take the decision of relocating here or there is not the same as the information used to decide where to locate a new concern. In particular, migrations within the same geographical market are likely to have more and better information about the sites than start-ups.

Notice, finally, that this theoretical framework says little about the sense of these differences in the determinants of industrial (re)location. This means that a priori we cannot predict which neoclassical, behavioural and institutional factors will (or will not) affect location and relocation decisions. It is also not clear whether it is different factors that affect locations and relocations or it is the same factors that affect both processes but with different intensity. Consequently, it seems that a sensible empirical strategy is to "let the data speak" and use the same vector of explanatory variables for the rate of both start-ups and relocations.

3.2 Empirical studies

Recent surveys by Blonigen (2005) and Arauzo and Manjón (2007) provide an excellent overview of the empirical literature on industrial location. However, this evidence basically refers to new concerns. Inferences from relocation data are much less common, especially those obtained conditioning on a set of explanatory variables. Mariotti (2005) reviews in detail the firm relocation literature since WWII until the late 1990s and find indeed that descriptive statistical methods prevail —see, however, Moses and Williamson (1967) and Erickson and Wasylenko (1980). Interesting as this research might be, it does not analyse partial or marginal effects.

As for the studies providing sounded econometric evidence, we have already mentioned the paper of Holl (2004a) because she resorts to CDM. However, there are a number of investigations on relocation that do not use CDM but DCM. In this literature we can further distinguish between those papers that are interested in the decision of "whether to relocate" (proxied by binary and ordered outcomes) and those that are interested in the decision of "where to locate" (proxied by nominal outcomes). Among the former we can mention Cooke (1983), van Dijk and Pellenbarg (2000), Brouwer et al. (2004) and Lee (2004). Among the latter we can mention Baudewyns et al. (2000) and Strauss-Kahn and Vives (2005). Another major difference between these studies comes from the data sources: Lee (2004) examines census data (as Holl (2004a) and this paper do) while the rest examine survey data.[14]

Cooke's (1983) study derives comparative statics from a model of intra-metropolitan firm relocation and empirically tests some of them using a probit specification. Data comes from a survey of manufacturing firms of Cincinnati that relocated 147 plants between 1971 and 1975-76, although the final sample only consists of 24 observations. The dependent variable is an indicator with value one if the distance between the initial and final location is more than three miles from the Central Business District. The explanatory variables are changes in demand and transport costs (multiplied by the change in the initial plant size, transport costs and agglomeration economies) and the "land intensity of technological change" (an indicator variable for the existence of technological change multiplied by "the change in land area of the plant in acres, as a result of the move").

---

[14] The statistical implications of using these distinct sources are worth noting. First, there is the issue of sample representativeness, apparent for example in Cooke (1983) and Strauss-Kahn and Vives (2005). Inferences are still valid conditioning on the sample, though authors do not always make this point clear enough. Second, as Barkley and

Also using survey data, Baudewyns et al. (2000) analyse the effects that better public infrastructures have on the (de)location decisions of Belgian firms from the city of Brussels (in the period 1981 to 1991) and the region of Wallonia (in the period 1990 to 1994) using a conditional logit model. Agglomeration economies and wage levels were also included among the regressors. Data come from STRATED for Brussels and Dun & Bradstreet for Wallonia. As for Van Dijk and Pellenbarg (2000), they analyse the stated preference of Dutch firms with regard to migration (measured as the propensity to move within the next two years reported in a questionnaire launched in 1995-1996) using ordered logit and probit models. Thanks to the detailed information they collected from the questionnaire these authors are able to distinguish between internal factors (organisational structure, financial reserves, size, etc.), external factors (labour market characteristics, government policies, general economic conditions, etc.) and location factors (occupancy characteristics, accessibility, distance to suppliers and markets, etc.) as determinants of relocation decisions.

While these two papers analyse firms from the same country, Brouwer et al. (2004) and Strauss-Kahn and Vives (2005) analyse firms from different countries. In particular, Brouwer et al. (2004) analyse a sample of large firms (more than 200 employees) from twenty-one countries. The dependent variable is binary ("In the 1999 [Cranet] survey, respondents were asked whether their firms have relocated in the last three years"), the econometric specification is a logit model and the main covariates are age, size, sector, size of the market and the region where the firm is located. As for Strauss-Kahn and Vives (2005), they use a nested logit to discern between the decision of "where to locate" and that of "whether to relocate". The explanatory variables are wages, population, sectorial distribution of employment, measures of agglomeration economies, corporate taxes and accessibility to airports. Data come from Dun & Bradstreet and official statistics. Also, two distinct features are worth mentioning about this paper: it presents a supportive theoretical model and the unit of analysis is US headquarters (from both US and non-US companies) rather than firms or plants.

McNamara (1994) show, stated preferences may be a poor proxy for location decisions. Inferences from survey data should therefore be interpreted with care.

Lastly, Lee (2004) aims to assess the impact of state development incentives on the decision to (re)locate. He uses a multinomial logit model in which the dependent variable takes three values: entry, relocation and exit. Rather than analysing the parameter estimates, however, in a second stage he evaluates the predicted probabilities for shutting down and relocating in a scenario of change vs. maintenance of the incentive policy. Moreover, he uses the same "method of recycled predictions" in a probit model for the decision to relocate within or away from the state. The explanatory variables consist of a vector of dummies for each incentive program as well as firm- and location-specific variables. Data covers the universe of US manufacturing establishments between 1972 and 1992, although he restricts the sample to multi-plant firms that relocated production (defined as the opening of a new plant more than 50 miles from the county of the original location, producing in the same four-digit industry and implying a reduction of more than 50% of the total employment in the original location).

4. Data and results

4.1 The data-base

Table 1 reports the definition and descriptive statistics of the dependent and explanatory variables used in this study. The data-base covers 946 municipalities in the period 2001 to 2004. However, Table 1 shows that there are some (random) missing data in the explanatory variables that make our final sample an unbalanced panel. We found no clear pattern in these missing values, i.e. they are not concentrated in, for example, small villages or in a particular province. It was consequently judged unnecessary to implement corrections for sample selection bias.

[Insert Table 1 about here]

The dependent variables in Panel A of Table 1 were calculated from the information available in the "Registry of Industrial Establishments of Catalonia". This administrative register contains all the establishments located in Catalan municipalities. However, since 2001 it provides (upon request) separate data for new and relocated establishments. To make the sample sectorially homogeneous we selected all establishments in codes 12 to 36 of the NACE-93 classification of the European Union. We have therefore limited the study to manufacturing establishments (plants).

Information regarding the explanatory variables comes basically from the data-base constructed by Trullén and Boix (2005). This was complemented with data from the Catalan Institute of Statistics. As Panel B of Table 1 shows, we have no variables related to the behavioural theory and dummies concerning the administrative and spatial organization of Catalonia are the only measures related to the institutional theory.

The fact that most regressors refer to the neoclassical theory may raise concerns about the omission of relevant variables. However, it is very difficult to asses the statistical relevance of these concerns in non-linear models like the ones used in this study. The fact that no empirical evidence exists showing the relative importance of each theory in location decisions does not help either. In any case, given that our unit of analysis is the municipality and not the firm, the potential biases caused by the lack of behavioural factors should be small. Also, as Pellenbarg et al. (2002a, 2002b) argue, behavioural theory is mostly a theory of entrepreneurship, whereas neoclassical and institutional theories appear to explain better the (re)location decisions of small and large firms respectively. We have no information on the number of entrepreneurs in our data-set, but we do know the size of the establishments: 65 per cent of them have more than three employees. It seems therefore that there is little risk of misspecification in our vector of explanatory variables.

The lack of behavioural factors contrasts however with the richness of information on certain institutional (metropolitan areas) and neoclassical (location economies, human capital and transport infrastructure) factors. To facilitate the interpretation of results, in Tables 2 to 5 we report Wald tests for these sets of coefficients rather than the estimated coefficients of each variable. “Wald Test L.E.”, for example, stands for the Wald test for the joint significance of the coefficients associated with the seventeen measures of location economies. The same goes for “Wald Test H.C.” (human capital), “Wald Test Infra.” (transport infrastructures) and “Wald Test Met.” (metropolitan areas).

Finally, it may be worth mentioning that there are two measures of urbanization and location economies in the data set. One is based on the total area of the municipality and the other on the area of the municipality defined as urban land. We opted for reporting results based on urban land because this seems a more accurate measure of the spatial area where economic activity occurs. However, results did not change substantially when using one or the other, the exception being the coefficients of density and (dis)urbanisation economies (these were not statistically significant when using measures based on the area of the municipality).

4.2 Estimates

In this section we present results from the estimation of count data models for the expected number of new and relocated industrial establishments created in Catalonia in the period 2001 to 2004, conditional to the same set of explanatory factors (summarised in Panel B of Table 1). In Tables 2 to 4 we present results under alternative assumptions on the exposure period, whereas in Tables 5 and 6 we address the potential dependence between start-ups and relocations. Besides a number of statistics (Wald and Goodness-of-fit tests, AIC, etc.), we report partial or marginal effects. We do not report coefficient estimates because our interest here is to empirically determine whether the opening of new and relocated establishments are processes driven by either different factors or the same factors but with different intensity. We are therefore interested in changes in the conditional mean of the dependent variables due to changes in these factors.[15]

Before presenting these inferences, however, it would be worth testing for equidispersion and the equality of the means of the dependent variables. After all, there is little point in the econometric design discussed in Section 2 unless we can reject these null hypotheses. To test for equidispersion in cross-section data we use an LR test based on the fact that the NBM reduces to the Poisson regression model if $\alpha = 0$. To test for overdispersion in panel data we use an auxiliary log-log regression between the variance of the

Pearson residuals from the pooled Poisson regression used to test for period-independence and the predicted values of the same model —see Hausman et al. (1984) for details on this procedure. All our specifications soundly rejected equidispersion. To test for the equality of means we use a classical conditional test and an unconditional test recently proposed by Krishnamoorthy and Thomson (2004).[16] We run the tests for yearly and pooled data, always finding that the null hypothesis was rejected. With these results, we can now confidently turn to analyse the statistical output from our models.

Irrespective of the model and data structures that we use, density, (dis)urbanisation economies and industrial diversity are all statistically significant variables with the expected sign (see e.g. Arauzo 2005 and Holl 2004a, 2004b and 2004c). The negative sign of entrepreneurship is at first sight surprising, but it may simply be reflecting the fact that entrants are mostly small firms competing in the same local markets as existent entrepreneurs. Also, institutional factors and industry shares tend to have a positive and significant effect on the expected number of start-ups and relocations.

[Insert Table 2 about here]

Although these general results remain largely unaltered across alternative specifications, we can observe differences in the estimated value of the partial effects and in the tests for the joint significance of location economies, human capital, transport infrastructure and metropolitan-dummy variables. It is therefore necessary to decide which is our preferred model for making these inferences. In this respect, estimates from the 2001 cross section in Table 3 present the highest likelihood and lowest penalized likelihood values. In particular, allowing for overdispersion and discrete heterogeneity (practically) provide the best fit. In addition, "LR Inflated" and "Vuong" tests reported in the last rows of Table 3 further support ZINBM over the other models. It seems therefore that in this case using a more

---

[15] For continuous covariates, partial effects are computed at their sample means; for dummies, they are calculated as the difference in the prediction of the dependent variable associated with the 0-to-1 change in the covariate.

[16] A Fortram program to compute the unconditional test is available at the web page of professor Krishnamoorthy.

parameterised model pays off. We shall consequently focus on the results from this model using 2001 data to analyse partial effects and Wald tests.

[Insert Table 3 about here]

The broad picture of neoclassical significant factors is essentially the same as described above, but neither institutional factors nor industry shares are now statistically significant. In any case, an interesting pattern emerges when comparing the partial effects on start-ups and relocations: values tend to be higher for the former than for the latter. This means that, although the determinants of both processes are basically the same, their impact differs, sometimes considerably. Also, these differences are consistent with the idea that the information on which location and relocation decisions hinge differs (Pellenbarg et al. 2002a, 2002b; Lee 2006). Lastly, Wald tests indicate that neither transport infrastructures nor "metropolitanity" are significant determinants of the frequency of strictly-new and relocated establishments. On the other hand, human capital and, in the case of start-ups, location economies matter.

[Insert Table 4 about here]

However, two caveats apply to these results. The first is whether they are robust to alternative ways of controlling for unobserved heterogeneity, i.e. whether panel data estimates make a difference. In light of the identification and convergence problems we faced when computing the FE estimator, only RE estimates can effectively be used for this purpose and these should be interpreted with care given the high values of the Hausman test (statistically significant for start-ups). Still, Table 4 shows that both Poisson and Negative Binomial RE estimates present similar statistical significances with respect to the neoclassical factors and opposite statistical significances with respect to institutional factors. However, the fit in these specifications is much worse than that of the 2001-ZINBM, with values of the AIC that double or triple those reported in Table 3.

[Insert Table 5 about here]

The second caveat arises from the signs of misspecification that the "GoF Test" shows. As discussed in Section 2, one possible source of misspecification is the failure of the independence assumption. Estimates reported in Tables 5 and 6 suggest that this assumption does not hold for our data, thus calling into question our previous conclusions. This is particularly true for the results reported in Table 5 using cumulative values of the dependent variables and period means as explanatory variables. However, since this specification still provides significant values for the "GoF Test", the appropriate comparison should be made with the results reported in Table 6. More specifically, we should compare results from 2001-ZINBM with those from ZIPM-ZINBM in Table 6.[17]

[Insert Table 6 about here]

Judging from these results the potential specification error affects mostly to the relocations' determinants, whereas for start-ups we find essentially the same significant variables and signs. Also, there exists a positive relation at the municipality level between the number of relocations (start-ups) and the number of start-ups (relocations). That is, the "average response" in the number of start-ups and relocations in a municipality to a variation in, respectively, the number of previous relocations and start-ups is always positive. However, this "signalling" or "vertical-integration" effect, which may also be interpreted as evidence of some kind of agglomeration economies, is not symmetric. Relocations seem to have a stronger effect on start-ups than vice versa. In fact, the partial effect of start-ups on relocations is very small. Once again, this reinforces the idea that relocating firms have more and better information on the potential sites.

4. Conclusions

[17] Since we cannot reject the null that $\alpha = 0$ when estimating the ZINBM, estimates from ZINBM in Table 6 are actually the same as those from ZIPM. Moreover, the negligible value of the estimate of α precluded the calculation of the "Gof Test" for ZINBM. It is also interesting to note that, although the fit of this specification is worse than that found when using the 2001 cross-section without interrelations (reported in Table 3), it is better than that found when using the 2002 cross-section without interrelations (not reported). In particular, the AIC and the Goodness-of-fit Test for ZINBM were 799.03 and 9.97 for start-ups and 402.15 and 6.06 for relocations.

This paper analyses the determinants of industrial location distinguishing between start-ups and relocations. We discuss several econometric strategies based upon count regression models for cross-section and panel data, thus assessing the robustness of our conclusions to alternative distributional assumptions about the data generating process. Interestingly, in this framework we can test not only for the distinct characteristics that make a municipality more attractive to start-ups and relocations but also for the existence of a relationship between these events.

We provide evidence that the location of new industrial establishments and the relocation of extant industrial establishments are driven by similar stochastic processes and determinants. However, the weight of these determinants, measured by the partial or marginal effect, differs between start-ups and relocations. We also find that locations and relocations are positively interrelated, although strictly-new establishments react on average more favourably to the presence of relocated establishments in the same municipality than relocated establishments to the presence of strictly-new establishments.

From an economic policy viewpoint, results indicate that public programs aiming to attract new businesses should pay attention to these issues. In addition to this general conclusion, however, the proposed econometric specifications enable us to address specific questions on the factors associated with the public sector (infrastructures, human capital, etc.), the private sector (e.g. agglomeration and location economies) and/or the space (distance to capitals, territorial dummies, etc.) that affect the expected number of new and relocated establishments per municipality. In the case of Catalonia, for example, it seems that certain public factors are not contributing to the likelihood functions.

References

Andrews, D.W.K. (1988): “Chi-Square Diagnostic Tests for Econometric Models. Introduction and Applications”, *Journal of Econometrics* **37**: 135-156.

Arauzo, J.M. (2005): “Determinants of Industrial Location. An Application for Catalan Municipalities”, *Papers in Regional Science* **84**: 105-120.

Arauzo, J.M. and Manjón, M. (2004): “Firm Size and Geographical Aggregation: An Empirical Appraisal in Industrial Location”, *Small Business Economics* **22**: 299-312.

Arauzo, J.M. and Manjón, M. (2007): “Empirical Studies in Industrial Location: An Assessment of their Methods and Results”, mimeo.

Barkley, D.L. and McNamara, K.T. (1994): “Manufacturer’s Location Decisions: Do Surveys Provide Helpful Insights?” *International Regional Science Review* **17(1)**: 23-47.

Baudewyns, D.; Ben Ayad, M. and Sekkat, K. (2000): “Infrastructure publique et localisation des entreprises à Bruxelles et en Wallonie”, in M. Beine and F. Docquier (eds.), *La politique de développement local et l'infrastructure publique: Bruxelles et Wallonie*, Bruxelles.

Becker, R. and Henderson, V. (2000): “Effects of Air Quality Regulations on Polluting Industries”, *Journal of Political Economy* **108**: 379– 421.

Blonigen, B.A. (2005): “A Review of the Empirical Literature on FDI Determinants,” *Atlantic Economic Journal* **33**: 383-403.

Brouwer, A.E., Mariotti, I. and van Ommeren, J.N. (2004): “The Firm Relocation Decision: An Empirical Investigation”, *The Annals of Regional Science* **38**: 335– 347.

Cameron, A.C. and Trivedi, P.K (1998): *Regression Analysis of Count Data*, Cambridge University Press, Cambridge.

Cameron, A.C. and Trivedi, P.K (2001): “Essentials of Count Data Regression”, in B. H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.

Carlton, D. (1979): “Why New Firms Locate Where They Do: An Econometric Model”, in W. Wheaton (ed.), *Interregional Movements and Regional Growth*. The Urban Institute, Washington.

Cooke, T.W. (1983): “Testing a model of intraurban firm relocation”, *Journal of Urban Economics* **13**: 257-282.

van Dijk, J. Pellenbarg, P. H. (2000): "Firm Relocation Decisions in the Netherlands: An Ordered Logit Approach", *Papers in Regional Science* **79**: 191-219.

Erickson, R.A. and Wasylenko, M. (1980): "Firm Relocation and Site Selection in Suburban Municipalities", *Journal of Urban Economics* **8**: 69-85.

Figueiredo, O.; Guimarães, P. and Woodward, D. (2002): "Home-field Advantage: Location Decisions of Portuguese Entrepreneurs", *Journal of Urban Economics* **52**: 341-361.

Frenkel, A. (2001): "Why High-Technology Firms Choose to Locate in or near Metropolitan Areas", *Urban Studies* **38**: 1083-1101.

Guimarães P.; Figueiredo, O. and Woodward, D. (2000): "Agglomeration and the location of foreign direct investment in Portugal", *Journal of Urban Economics* **47**: 115–135.

Guimarães, P.; Figueiredo, O. and Woodward, D. (2003): "A Tractable Approach to the Firm Location Decision Problem", *Review of Economics and Statistics* **85(1)**: 201-204.

Guimarães, P.; Figueiredo, O. and Woodward, D. (2004): "Industrial Location Modeling: Extending the Random Utility Framework", *Journal of Regional Science* **44 (1)**: 1-20.

Hausman, J.A.; Hall, B.A. and Griliches, Z. (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship", *Econometrica* **52 (4)**: 909-938.

Hayter, R. (1997): *The dynamics of industrial location. The factory, the firm and the production system*, New York: Wiley.

Holl, A. (2004a): "Start-ups and Relocations: Manufacturing Plant Location in Portugal", *Papers in Rgional Science* **83**: 649-668.

Holl, A. (2004b): "Transport Infrastructure, Agglomeration Economies, and Firm Birth. Empirical Evidence from Portugal", *Journal of Regional Science* **44**: 693-712.

Holl, A. (2004c): "Manufacturing Location and Impacts of Road Transport Infrastructure: Empirical Evidence from Spain", *Regional Science and Urban Economics* **34**: 341-363.

Krishnamoorthy, K. and Thomson, J. (2004): "A More Powerful Test for Comparing Two Poisson Means", *Journal of Statistical Planning and Inference* **119**: 23-35.

Lee, Y. (2004): "Geographic Redistribution of US Manufacturing and the Role of State Development Policy", Working Paper 04-15, Federal Reserve Bank of Cleveland.

Lee, Y. (2006): “Relocation Patterns in US Manufacturing”, Working Paper 06-24, Federal Reserve Bank of Cleveland.

List, J.A. and McHone, W.W. (2000): “Measuring the Effects of Air Quality Regulations on “Dirty” Firm Births: Evidence from the Neo and Mature-Regulatory Periods”, *Papers in Regional Science* **79**: 177-190.

Mariotti, I. (2005): *Firm relocation and regional policy*, Utrecht / Groningen: Department of Spatial Sciences (University of Groningen), Netherlands Geographical Studies 331.

Moses, L. and Williamson, H.F. (1967): “The Location of Economic Activity in Cities”, *American Economic Review* **57**: 211-222.

Mullahy, J. (1997): “Heterogeneity, Excess Zeros, and the Structure of Count Data Models”, *Journal of Applied Econometrics* **12**: 337-350.

Papke, L. (1991): “Interstate Business Tax Differentials and New Firm Location”, *Journal of Public Economics* **45**: 47-68.

Pellenbarg, P.H., van Wissen, L.J.G. and van Dijk, J. (2002a): “Firm Relocation: State of the Art and Research Prospects”, SOM Research Report 02D31, University of Groningen.

Pellenbarg, P.H., van Wissen, L.J.G. and van Dijk, J (2002b): “Firm Migration”, in P. McCann (ed.), *Industrial Location Economics*. Edward Elgar Publishing, Cheltenham.

Sin, C.-Y. and White, H. (1996): “Information criteria for selecting possibly misspecified parametric models”, *Journal of Econometrics* **71**: 207-225.

Stam, E. (2007): “Why Butterflies Don’t Leave: Locational Behavior of Entrepreneurial Firms”, *Economic Geography* **83**: 27-50.

Strauss-Kahn, V. and Vives, X. (2005): “Why and Where Do Headquarters Move?”, Discussion Paper No. 5070, Centre for Economic Policy Research.

Trullén, J. and Boix, R. (2005): *Indicadors 2005*, Diputació de Barcelona i Universitat Autònoma de Barcelona.

Figure 1: Frequency distribution of Start-ups, Relocations and Poisson-generated Random Variable with Unitary Mean.

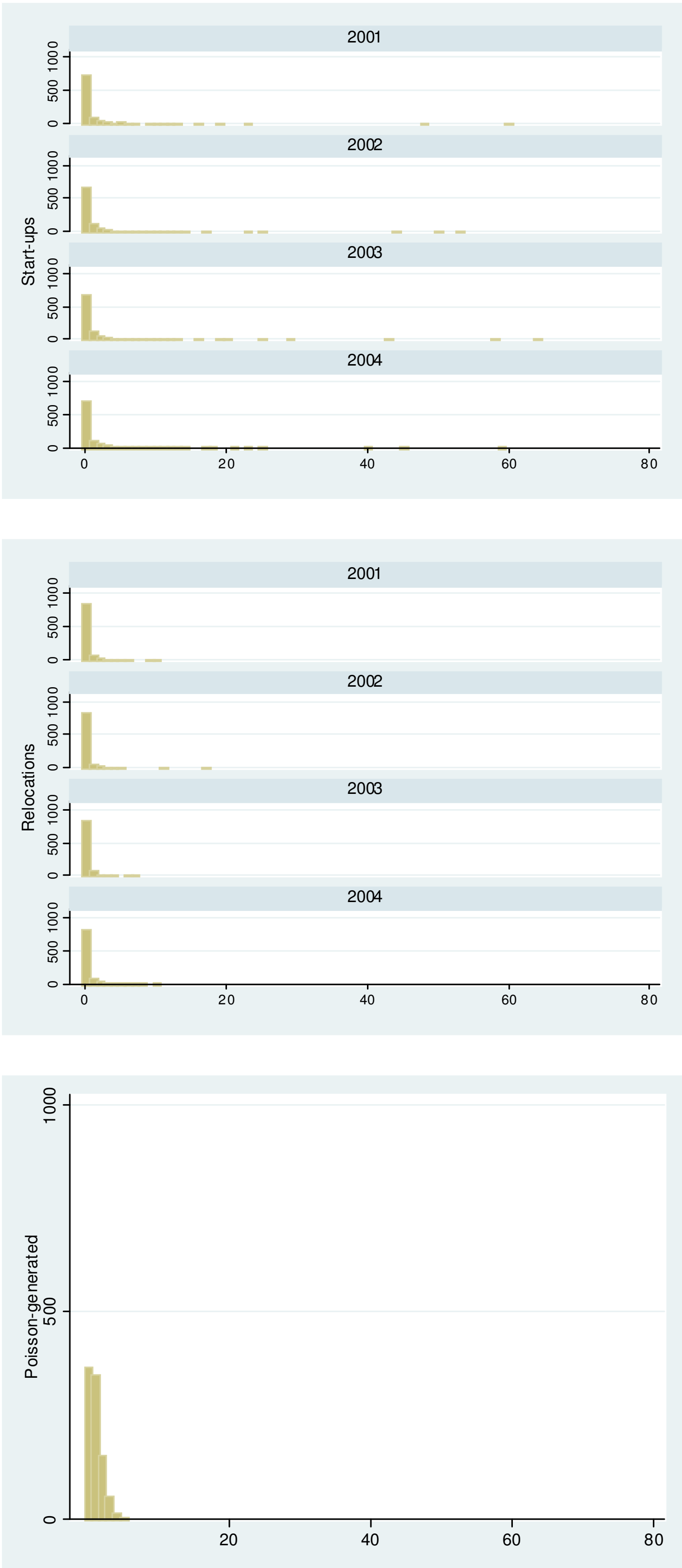

Table 1: Descriptive statistics (pooled sample).

	Period	Obs.	Mean	S. D.	Min.	Max.
Panel A: Dependent Variables						
Start-ups	2001-2004	3784	0.93	3.55	0	64
Relocations	2001-2004	3784	0.19	0.75	0	17
Panel B: Explanatory Variables						
B.1 Neoclassical approach						
Population density (*DENS*)	2001-2004	3530	40.43	100.13	0.74	3220
Urbanisation economies: Workers per km^2 (*URB*)	2001-2004	3530	10.47	33.73	0	980
Location Economies: Workers in sector "i" (*i = 1, …, 17*) per km^2.	2001-2004	3648	-	-	-	-
Industrial diversity: Inverse of the H-H index of diversity of industrial jobs (*DIV*)	2001-2004	3758	3.57	1.41	1	7.43
Industry share: % of jobs in the industrial sector (*INDS*)	2001	3784	0.22	0.11	0	0.60
Services share: % of jobs in the services sector (*SERS*)	2001-2004	3758	0.48	0.24	0	1
Entrepreneurship: % of jobs declared to be entrepreneurs (*ENTREP*)	2001	3784	0.26	0.11	0.02	0.69
Commuting: Mobility ratio (*COMM*)	2001	3780	1.79	17.30	0	521.07
Human capital: % of population working in science and technology	2001	3784	0.19	0.05	0.06	0.47
Human capital: % of population with a university degree	2001	3784	0.14	0.05	0.01	0.5
Human capital: Years of education, average for population over 25 years old	2001	3784	8.5	1.01	4.23	11.99
Infrastructures: Average travelling time to the capitals of province	2001	3768	87.38	23.30	56	190
Infrastructures: Average travelling time to nearest airport	2001	3768	49.07	32.98	0	190
Infrastructures: Average travelling time to nearest port	2001	3768	62.44	32.99	0	197
Infrastructures: Dummy for train station	2001-2004	3780	0.10	0.30	0	1
B.2 Institutional approach						
Dummy for the province of Barcelona (*PROV1*)	2001-2004	3784	0.32	0.46	0	1
Dummy for the province of Girona (*PROV2*)	2001-2004	3784	0.23	0.42	0	1
Dummy for the province of Lleida (*PROV3*)	2001-2004	3784	0.24	0.42	0	1
Dummy for capital of "*comarca*" (*CAPC*)	2001-2004	3784	0.04	0.20	0	1
Dummy for coastal municipality (*COAST*)	2001-2004	3784	0.07	0.26	0	1
Dummy for metropolitan area of Barcelona	2001-2004	3784	0.23	0.42	0	1
Dummy for metropolitan area of Girona	2001-2004	3784	0.07	0.26	0	1
Dummy for metropolitan area of Lleida	2001-2004	3784	0.06	0.24	0	1
Dummy for metropolitan area of Manresa	2001-2004	3784	0.03	0.17	0	1
Dummy for metropolitan area of Tarragona	2001-2004	3784	0.07	0.26	0	1

Source: "Registry of Industrial Establishments of Catalonia" (dependent variables), Trullén and Boix (2005) and Catalan Institute of Statistics (explanatory variables).

Note: To save space, details on the measures of location economies are not reported. The fourth province of Catalonia (not reported) is Tarragona. A "*comarca*" is a territorial unit formed by adjacent municipalities. There are 11 "*comarques*" in the province of Barcelona, 8 in Girona, 12 in Lleida and 10 in Tarragona.

Table 2: Determinants of (cumulative 2001-2004) Start-ups and Relocations.

	Start-ups			Relocations		
	NBM	**ZIPM**	**ZINBMZINBM**	**NBM**	**ZIPM**	**ZINBMZINBM**
DENS	-0.0134 (0.0026)***	-0.0130 (0.0016)***	-0.0162 (0.0035)***	-0.0031 (0.0007)***	-0.0045 (0.0012)***	-0.0059 (0.0016)***
URB	0.0358 (0.0080)***	0.0655 (0.0054)***	0.0536 (0.0120)***	0.0098 (0.0022)***	0.0232 (0.0042)***	0.0229 (0.0055)***
URBA^2	-7.2×10^{-5} (10^{-6})***	-0.0001 (10^{-6})***	-9.9×10^{-5} (-2.0×10^{-5})***	-1.6×10^{-5} (-8.8×10^{-5})***	-4.3×10^{-5} (10^{-6})***	-4.6×10^{-5} (10^{-6})***
DIV	0.3568 (0.0386)***	0.1939 (0.0298)***	0.3650 (0.0513)***	0.0810 (0.0123)***	0.0800 (0.0211)***	0.1016 (0.0243)***
INDS	0.9059 (0.6404)	3.3522 (0.6260)***	1.5654 (0.9699)*	0.3847 (0.2037)*	1.5431 (0.4620)***	1.0238 (0.4870)**
SERS	-0.3773 (0.2607)	1.0607 (0.2718)***	-0.1377 (0.4073)	-0.0364 (0.0918)	0.0298 (0.1923)	0.0280 (0.2133)
ENTREP	-3.6275 (0.7654)***	-5.6298 (0.8339)***	-3.4208 (1.1508)***	-0.8115 (0.2588)***	-1.3407 (0.5794)**	-1.0641 (0.5984)*
COMM	0.0011 (0.0014)	-3.7×10^{-5} (0.0006)	0.0013 (0.0018)	0.0001 (0.0003)	0.0004 (0.0004)	0.0002 (0.0006)
PROV1	0.8743 (0.2362)***	1.6899 (0.2589)***	1.2942 (0.3419)***	0.2566 (0.1011)**	0.6602 (0.2213)***	0.6687 (0.2523)***
PROV2	0.6197 (0.3526)*	1.1193 (0.3422)***	1.0036 (0.5271)*	0.3859 (0.1778)**	0.9373 (0.3595)***	0.8496 (0.3895)**
PROV3	1.2683 (0.4502)***	2.5751 (0.4996)***	1.8946 (0.6782)***	0.4107 (0.2051)**	0.8809 (0.3976)**	0.9258 (0.4718)**
CAPCOM	0.2675 (0.2580)	1.2935 (0.2327)***	0.5235 (0.3834)	0.2195 (0.1155)*	0.37701 (0.1699)**	0.4201 (0.2278)*
COAST	0.8961 (0.2896)***	0.8067 (0.1427)***	0.8834 (0.3268)***	0.1884 (0.0801)**	0.4020 (0.1239)***	0.2563 (0.1332)*
Wald Test L.E.	62.17***	472.30***	61.91***	63.36***	86.47***	61.18***
Wald Test H.C.	8.01**	31.66***	6.83*	4.54	6.71*	5.47
Wald Test Infra.	19.32***	64.74***	17.01***	20.20***	24.74***	19.64***
Wald Test Met.	2.97	76.32***	4.79	13.12**	17.02***	8.15
Log L	-1357.13	-1743.93	-1330.92	-680.23	-682.52	-650.43
AIC	1445.13	1833.93	1422.91	768.23	772.52	742.43
LR Joint Test	897.08***	5663.09***	648.78***	570.79***	667.71***	292.43***
GoF Test	16.81***	45.23***	28.18***	26.43***	19.70***	9.62**
LR Inflated Test		826.03***			64.18***	
Vuong Test		3.67***	3.20***		4.05***	3.58***

Note: Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors are in brackets. The symbols ***, ** and * denote significance at 1%, 5% and 10% levels, respectively (907 observations). See sections 2 and 4 for definitions of the reported statistics.

<u>Table 3</u>: Determinants of (2001) Start-ups and Relocations.

	Start-ups			Relocations		
	NBM	**ZIPM**	**ZINBM**	**NBM**	**ZIPM**	**ZINBM**
DENS	-0.0015	-0.0024	-0.0028	-0.0002	-0.0004	-0.0005
	(0.0004)***	(0.0007)***	(0.0010)***	(0.0001)*	(0.0002)*	(0.0003)
URB	0.0078	0.0204	0.0193	0.0010	0.0025	0.0027
	(0.0019)***	(0.0046)***	(0.0054)***	(0.0004)**	(0.0011)**	(0.0013)**
URBA^2	-2.2×10^{-5}	-4.8×10^{-5}	-5.6×10^{-5}	-6.3×10^{-6}	-1.5×10^{-5}	-1.5×10^{-5}
	(0.0001)***	(10^{-6})***	(10^{-6})***	(10^{-6})***	(10^{-6})**	(10^{-6})**
DIV	0.0355	0.0237	0.0475	0.0046	0.0060	0.0074
	(0.0084)***	(0.0131)*	(0.0165)***	(0.0025)*	(0.0038)	(0.0048)
INDS	0.1887	1.1434	0.5307	0.0411	0.1489	0.1444
	(0.1266)	(0.3255)***	(0.3527)	(0.0310)	(0.0949)	(0.1037)
SERS	-0.0494	0.0825	-0.0018	0.0176	0.0787	0.0805
	(0.0540)	(0.1285)	(0.1517)	(0.0133)	(0.0468)*	(0.0530)
ENTREP	-0.4438	-0.8330	-0.5005	-0.0071	0.1212	0.1214
	(0.1954)**	(0.4257)**	(0.4982)	(0.0344)	(0.1107)	(0.1203)
COMM	-0.0008	-0.0023	-0.0021	0.0002	0.0005	0.0005
	(0.0014)	(0.0024)	(0.0032)	(0.0002)	(0.0005)	(0.0006)
PROV1	-0.0077	0.1213	0.0287	0.01450	0.0463	0.0494
	(0.0382)	(0.0932)	(0.0984)	(0.0149)	(0.0424)	(0.0486)
PROV2	-0.0283	-0.0856	-0.0422	0.0129	0.0392	0.0433
	(0.0379)	(0.0800)	(0.0991)	(0.0173)	(0.0489)	(0.0549)
PROV3	0.0144	0.1931	0.0231	0.0193	0.0721	0.0833
	(0.0576)	(0.1606)	(0.1371)	(0.0240)	(0.0785)	(0.0950)
CAPCOM	0.0799	0.3470	0.2117	0.0314	0.0681	0.0800
	(0.0597)	(0.1352)***	(0.1459)	(0.0275)	(0.0571)	(0.0718)
COAST	0.0762	0.2497	0.1334	0.0022	0.0150	0.0051
	(0.0426)*	(0.0849)***	(0.0853)	(0.0062)	(0.0169)	(0.0162)
Wald Test L.E.	38.94***	123.26***	37.83***	18.96	29.54**	22.02
Wald Test H.C.	7.35*	14.02***	8.42*	10.01**	10.46**	8.79**
Wald Test Infra.	3.92	10.17*	1.87	4.52	5.17	5.13
Wald Test Met.	4.09	16.64***	4.64	10.49**	8.52	8.02
Log L	-596.51	-612.07	-576.52	-237.77	-228.19	-227.24
AIC	684.51	702.07	668.52	325.77	318.19	319.24
LR Joint Test	522.26***	873.75***	266.93***	264.13***	180.02***	131.42***
GoF Test	8.88**	43.28***	28.52***	12.65***	15.80***	11.25**
LR Inflated Test		71.09***			1.88*	
Vuong Test		3.56***	3.33***		2.07**	1.96**

<u>Note</u>: Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors are in brackets. The symbols ***, ** and * denote significance at 1%, 5% and 10% levels, respectively (816 observations). See sections 2 and 4 for definitions of the reported statistics.

Table 4: Determinants of Start-ups and Relocations (2001-2004 panel).

	Start-ups				Relocations			
	POIS-FE	**POIS-RE**	**NB-FE**	**NB-RE**	**POIS-FE**	**POIS-RE**	**NB-FE**	**NB-RE**
DENS	0.0126	-0.0081	-0.0180	-0.0080	0.0009	-0.0117	-0.0093	-0.0112
	(0.0070)*	(0.0027)***	(0.0069)***	(0.0026)***	(0.0111)	(0.0035)***	(0.0102)	(0.0039)***
URB	-0.0072	0.0314	-0.0046	0.0309	-0.0149	0.0454	0.0160	0.0374
	(0.0093)	(0.0069)***	(0.0111)	(0.0070)***	(0.0238)	(0.0106)***	(0.0267)	(0.0114)***
URBA^2	1.4×10^{-5}	-3.8×10^{-5}	1.7×10^{-5}	-3.7×10^{-5}	1.5×10^{-5}	-4.9×10^{-5}	-9.9×10^{-5}	-3.8×10^{-5}
	(10^{-6})	$(10^{-6})^{***}$	(2.0×10^{-5})	$(10^{-6})^{***}$	(2.0×10^{-5})	$(10^{-6})^{***}$	(8.0×10^{-5})	$(10^{-6})^{***}$
DIV	0.0606	0.3204	0.0967	0.3369	0.1537	0.4006	0.0847	0.3707
	(0.0712)	(0.0374)***	(0.0752)	(0.0373)***	(0.1602)	(0.0571)***	(0.1463)	(0.0607)***
INDS	-	0.7424	-	0.4890	-	1.7982	-	1.2159
		(0.7668)		(0.7615)		(1.0894)*		(1.1659)
SERS	-0.3202	-0.1635	-0.3461	-0.1884	2.5295	0.2647	-0.9719	0.2566
	(0.6691)	(0.2754)	(0.6792)	(0.2800)	(1.3975)*	(0.4461)	(1.1726)	(0.4676)
ENTREP	-	-5.3736	-	-5.4259	-	-4.7148	-	-5.3543
		(0.9472)***		(0.9678)***		(1.4274)***		(1.5052)***
COMM	-	0.0014	-	0.0014	-	0.0013	-	0.0013
		(0.0018)		(0.0017)		(0.0019)		(0.0021)
PROV1	-	1.0417	-	1.1046	-	1.0945	-	0.9934
		(0.2033)***		(0.2077)***		(0.3411)***		(0.3502)***
PROV2	-	0.7248	-	0.7029	-	1.3695	-	1.2884
		(0.3002)**		(0.3042)**		(0.4114)***		(0.4295)***
PROV3	-	1.2970	-	1.2322	-	1.3899	-	1.2357
		(0.2900)***		(0.2889)***		(0.4665)***		(0.4819)***
CAPCOM	-	0.7745	-	0.7526	-	1.0369	-	1.1271
		(0.2412)***		(0.2325)***		(0.2949)***		(0.3215)***
COAST	-	0.6307	-	0.6383	-	0.6602	-	0.4914
		(0.1794)***		(0.1796)***		(0.2290)***		(0.2574)*
Wald Test L.E.	17.25	29.31**	40.02***	35.45***	4.76	49.12***	13.88	35.97***
Wald Test H.C.	-	11.42***	-	13.29***	-	7.40*	-	4.89
Wald Test Infra.	-	23.81***	-	19.13***	-	22.34***	-	18.16***
Wald Test Met.	-	2.73	-	2.15	-	13.96**	-	10.72*
Log L	-1400.71	-2788.12	-1385.40	-2760.66	-568.66	-1261.65	-611.42	-1286.03
AIC	1444.71	2876.12	1431.40	2850.66	612.66	1349.65	657.42	1376.03
LR Joint Test	18.68	1103.82***	46.88***	1121.63***	10.15	568.53***	69.56***	450.77***
Hausman Test	66.63***		62.94***		15.80		26.81	

Note: POIS(NB)-FE and POIS(NB)-RE denote Poisson (Negative Binomial) Fixed and Random Effects, respectively. Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors are in brackets. The symbols ***, ** and * denote significance at 1%, 5% and 10% levels, respectively (1634 and 3499 observations for FE and RE estimates, respectively). See sections 2 and 4 for definitions of the reported statistics. Estimates from POIS-FE (Start-ups), NB-FE (Relocations) and NB-RE (Relocations) were obtained under weak convergence criteria.

<u>Table 5</u>: Determinants of (cumulative 2002-2004) Start-ups and Relocations (2001 Interrelations).

	Start-ups			**Relocations**		
	NBM	**ZIPM**	**ZINBM**	**NBM**	**ZIPM**	**ZINBM**
DENS	-0.0084 (0.0020)***	-0.0071 (0.0017)***	-0.0093 (0.0029)***	-0.0020 (0.0006)***	-0.0043 (0.0013)***	-0.0044 (0.0014)***
URB	0.0290 (0.0062)***	0.0570 (0.0060)***	0.0455 (0.0092)***	0.0075 (0.0019)***	0.0189 (0.0042)***	0.0167 (0.0047)***
URBA^2	-5.7×10^{-5} (10^{-6})***	-9.1×10^{-5} (10^{-6})***	-7.9×10^{-5} (10^{-6})***	-1.4×10^{-5} (10^{-6})***	-4.5×10^{-5} (10^{-6})***	-4.1×10^{-5} (10^{-6})***
DIV	0.2719 (0.0316)***	0.1977 (0.0328)***	0.2798 (0.0454)***	0.0636 (0.0108)***	0.0894 (0.0219)***	0.0836 (0.0220)***
INDS	0.4832 (0.5515)	0.1786 (0.6980)	0.8299 (0.8709)	0.3393 (0.1877)*	0.7189 (0.4372)	0.7203 (0.4375)
SERS	-0.1716 (0.2219)	-0.1315 (0.2872)	-0.0455 (0.3768)	-0.1000 (0.0821)	-0.4114 (0.2004)*	-0.2387 (0.1991)
ENTREP	-3.1042 (0.6494)***	-5.9744 (0.8973)***	-3.0042 (1.0443)***	-0.8069 (0.2427)***	-1.6911 (0.5802)***	-1.1645 (0.5665)**
COMM	0.0003 (0.0011)	-3.8×10^{-5} (0.0007)	0.0004 (0.0015)	0.0002 (0.0002)	0.0006 (0.0004)	0.0004 (0.0005)
PROV1	0.9765 (0.2315)**	1.4881 (0.2830)***	1.4924 (0.3506)***	0.1722 (0.0813)**	0.4134 (0.1951)**	0.4519 (0.2036)**
PROV2	0.7718 (0.3630)**	1.4739 (0.4432)***	1.1896 (0.5455)**	0.2767 (0.1413)**	0.8081 (0.3536)**	0.6604 (0.3304)**
PROV3	1.1938 (0.4110)	1.9828 (0.4793)***	1.7955 (0.6329)***	0.2663 (0.1565)*	0.4481 (0.3040)	0.4950 (0.3251)
CAPCOM	0.2571 (0.2252)	0.3283 (0.1726)*	0.4874 (0.3346)	0.1034 (0.0760)***	-0.0247 (0.0981)	0.1167 (0.1357)
COAST	0.7423 (0.2432)***	0.0293 (0.1164)	0.7230 (0.2750)***	0.1743 (0.0735)**	0.1691 (0.0999)*	0. 2061 (0.1140)*
REL	0.1638 (0.0410)***	0.2863 (0.0254)***	0.2535 (0.0550)***			
STU				0.0092 (0.0028)***	0.0168 (0.0035)***	0.0204 (0.0056)***
Wald Test L.E.	52.46***	230.00***	51.13***	54.64***	73.26***	53.21***
Wald Test H.C.	3.34	8.33**	2.39	2.39	4.95	3.71
Wald Test Infra.	19.13***	59.39***	16.49***	19.22***	12.73***	13.17***
Wald Test Met.	3.08	22.31***	6.20	8.00	7.22	4.27
Log L	-1225.41	-1356.66	-1195.63	-608.79	-592.33	-581.22
AIC	1315.41	1448.66	1289.63	698.79	592.33	675.22
LR Joint Test	860.34***	4449.75***	631.31***	518.03***	544.98***	273.75***
GoF Test	20.54***	52.59***	33.81***	17.52***	11.72***	7.39*
LR Inflated Test		322.07***			22.23***	
Vuong Test		5.30***	3.53***		3.71***	3.55***

<u>Note</u>: *REL* (*STU*) denotes 2001 relocations (start-ups). Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors are in brackets. The symbols ***, ** and * denote significance at 1%, 5% and 10% levels, respectively (906 observations). See sections 2 and 4 for definitions of the reported statistics.

Table 6: Determinants of 2002 Start-ups and Relocations (2001 Interrelations).

	Start-ups			Relocations		
	NBM	**ZIPM**	**ZINBM**	**NBM**	**ZIPM**	**ZINBM**
DENS	-0.0020	-0.0029	-0.0030	-0.0001	-0.0007	-0.0007
	(0.0006)***	(0.0008)***	(0.0010)***	(0.0001)	(0.0005)	(0.0005)
URB	0.0121	0.0198	0.0177	0.0029	0.0072	0.0072
	(0.0024)***	(0.0034)***	(0.0039)***	(0.0008)***	(0.0023)***	(0.0023)***
$URBA^2$	-3.0×10^{-5}	-4.7×10^{-5}	-4.8×10^{-5}	-1.7×10^{-5}	-2.2×10^{-5}	-2.2×10^{-5}
	$(10^{-6})^{***}$	$(10^{-6})^{***}$	$(10^{-6})^{***}$	$(10^{-6})^{***}$	(10^{-6})	(10^{-6})
DIV	0.0521	0.0488	0.0530	0.0080	0.0171	0.0171
	(0.0110)***	(0.0146)***	(0.0151)***	(0.0034)**	(0.0079)**	(0.0079)**
INDS	0.1002	0.0980	0.2315	0.0494	0.2996	0.2996
	(0.1946)	(0.3158)	(0.3096)	(0.0453)*	(0.1739)*	(0.1739)*
SERS	0.0202	0.0281	0.0255	-0.0043	-0.0644	-0.0644
	(0.0767)	(0.1352)	(0.1286)	(0.0173)	(0.0667)	(0.0667)
ENTREP	-0.8871	-1.8754	-1.0495	-0.1022	-0.1922	-0.1922
	(0.2870)***	(0.4842)***	(0.4465)***	(0.0732)	(0.2121)	(0.2121)
COMM	-0.0002	-0.0003	-0.0004	-0.0001	-0.0003	-0.0003
	(0.0002)	(0.0003)	(0.0004)	(0.0002)	(0.0006)	(0.0006)
PROV1	0.1559	0.1733	0.2288	0.0277	0.0668	0.0668
	(0.0649)**	(0.0891)**	(0.0977)**	(0.0229)	(0.0654)	(0.0654)
PROV2	-0.0255	-0.0099	-0.0117	0.0533	0.1755	0.1755
	(0.0736)	(0.1157)	(0.1159)	(0.0420)	(0.1400)	(0.1400)
PROV3	0.1145	0.1360	0.1351	0.0033	-0.0058	-0.0058
	(0.0949)	(0.1341)	(0.1366)	(0.0245)	(0.0719)	(0.0719)
CAPCOM	0.3106	0.2451	0.3874	0.0268	-0.0262	-0.0262
	(0.1311)**	(0.1090)**	(0.1735)**	(0.0253)	(0.0448)	(0.0448)
COAST	-0.0712	0.0023	0.0816	0.0105	0.0063	0.0063
	(0.0520)	(0.0525)	(0.0707)	(0.0118)	(0.0273)	(0.0273)
REL	0.0468	0.0722	0.0735			
	(0.0109)***	(0.0129)***	(0.0175)***			
STU				0.0014	0.0034	0.0034
				(0.0007)**	(0.0016)**	(0.0016)**
Wald Test L.E.	36.80***	78.65***	39.05***	14.57	21.71	21.71
Wald Test H.C.	8.87**	10.68**	9.21	0.69	2.59	2.59
Wald Test Infra.	5.45	7.03	4.83	10.30**	6.65	6.65
Wald Test Met.	2.82	7.55	4.87	6.49	5.15	5.15
Log L	-705.53	-713.78	-690.32	-314.43	-300.45	-300.45
AIC	795.53	805.78	784.32	404.43	392.45	394.45
LR Joint Test	665.81***	1221.56***	430.73***	269.23***	198.64***	134.98***
GoF Test	5.59	27.15***	8.02**	14.93***	3.26	-
LR Inflated Test		46.92***			0.00	
Vuong Test		2.89***	2.62***		2.61***	15.17***

Note: *REL* (*STU*) denotes 2001 relocations (start-ups). Marginal or partial effects of selected explanatory variables (see Panel B of Table 1) are reported. Standard errors are in brackets. The symbols ***, ** and * denote significance at 1%, 5% and 10% levels, respectively (876 observations). See sections 2 and 4 for definitions of the reported statistics.