STUDY CENTER
GERZENSEE

On Discrete Location Choice Models

Nils Herger

Working Paper 13.02

On Discrete Location Choice Models

February, 2013

Abstract

Within the context of the firm location choice problem, Guimarães *et al.* (2003) have shown that a Poisson count regression and a conditional logit model yield identical coefficient estimates. Yet, the corresponding interpretation differs since these discrete choice models reflect polar cases as regards the degree with which the different locations are similar. Schmidheiny and Brülhart (2011) have shown that these cases can be reconciled by adding a fixed outside option to the choice set and transforming the conditional logit into a nested logit framework. This gives rise to a dissimilarity parameter (λ) that equals 1 for the Poisson count regression (where locations are completely dissimilar) and 0 for the conditional logit model (where locations are completely similar). Though intermediate values are possible, the nested logit framework does not permit the dissimilarity parameter to be pinned down. We show that, with panel data and adopting a choice consistent normalisation, the fixed outside option can also be introduced into the Poisson count framework, from which the estimation of the dissimilarity parameter is relatively straightforward. The different location choice models are illustrated with an empirical application using cross-border acquisitions data.

JEL classification: C25, F23

Keywords: Conditional Logit Model, Location Choice, Nested Logit Model, Poisson Count Regression

1 Introduction

Econometric models drawing on the discrete choices that are revealed when, for example, firms decide to locate in a given place, provide a popular framework to study the effect of economic, political, or other factors on the geographical distribution of economic activities. Location choices can often be observed on a comprehensive basis since they manifest in such things as the establishment of new plants. For the case of foreign direct investment (FDI), for example, Devereux and Griffith (1998) uncover the extent to which corporate taxes influence the decision of US multinational enterprises (MNEs) to install local production capacity either in France, Germany, or the United Kingdom. Similarly, Buettner and Ruf (2007) analyse how corporate taxation affects the observed distribution of foreign plants affiliated to German multinationals, whilst Kim *et al.* (2003), Crozet *et al.* (2004), and Devereux *et al.* (2007) all study the role of agglomeration effects for FDI through the lens of individual location choices. All these studies account for the discrete nature of location choices by using an empirical framework of the binary or conditional logit class. Rather than contemplating the decisions of individual firms, a different strand of the literature draws on *aggregated* counts of such location choices. Considering again the example of FDI, Kessing *et al.* (2007), Herger *et al.* (2008), Hijzen *et al.* (2008), and Coeurdacier *et al.* (2009) all employ the number, or count, of cross-border acquisitions (CBAs) to uncover the determinants affecting the desire of MNEs to place economic activities in a given location. All these studies account for the discrete and non-negative nature of count data by employing regressions of the Poisson class.

The different approaches to analysing the empirical determinants to locate economic activities raise the question of the econometric and economic differences between the conditional logit approach with individual, and the Poisson count regression with aggregate location choices. Though these econometric models have been developed independently, they share some similarities. Specifically, for both cross-sectional and panel data, Guimarães *et al.* (2003) have shown that the conditional logit model and Poisson count regression give rise to *identical* coefficient estimates. This favours the usage of count data since the aggregation of, say, location choices entails a possibly dramatic reduction in the number of observations required for estimation. However, according to Schmidheiny and Brülhart (2011)—henceforth SB—it is nevertheless important to distinguish between alternative location choice models. In particular, though the coefficient estimates are identical, the results of the Poisson count regression and conditional logit model differ when re-expressing the results in terms of elasticities. The reason is that in the Poisson count regression the aggregate number of location choices is used as the dependent variable, which can change with the value of the regressors. Borrowing the terminology of SB, the Poisson count regression reflects a "positive sum world". Furthermore, when using count data, the locations are segmented in the sense that only the local conditions, but not those elsewhere, affect the dependent variable. Conversely, the conditional logit model reflects a "zero sum world" since individual location choices are the dependent variable whilst their total number is exogenously fixed. This gives rise to spillover effects in the sense that an additional choice of a given location is always offset by an equivalent reduction elsewhere.

SB show that adding a fixed outside option gives rise to a nested logit model that encompasses the polar cases embodied in the Poisson count regression and the conditional logit model. This may be intuitive since the main innovation of transforming the conditional logit into a nested logit model is to summarise (location) choices into groups, or nests, that can be more or less similar. The degree of segmentation between these groups manifests in the so-called dissimilarity parameter λ (sometimes also called the log-sum coefficient or inclusive value parameter). Within the context of location choices, SB show that the Poisson count regression implies that $\lambda = 1$, that is the elemental options are completely dissimilar, whilst the conditional logit model arises when $\lambda = 0$, that is elemental options are entirely similar.

Yet, λ can in principle adopt any intermediate value on $[0, 1]$. Unfortunately, the nested logit framework of SB does not permit them to pin down the empirical value of λ.

This paper contributes to the literature by suggesting a way to extract the dissimilarity parameter from aggregated count data within a Poisson regression framework. One reason why λ cannot be estimated in the SB framework is that their nested logit model is over-parameterised in terms of an overlap between the coefficient pertaining to the fixed outside option and the dissimilarity parameter that, hence, drops out of the likelihood function. It is well known that nested logit models suffer from an identification problem and, hence, require some suitable normalisation of the so-called scale parameters (Hunt, 2000; Ben Aktiva and Lerman, 1985; Hensher *et al.*, 2005, ch.13). This step, which is arguably often ignored in applied work, is crucial insofar as it gives rise to different versions of the nested logit model with different properties (Hunt, 2000; Hensher and Greene, 2005). Thereby, a minimal requirement would be that the normalisation is consistent with the basic principles that are thought to guide the (location) choices (Koppelman and Wen, 1998). For example, with a choice consistent normalisation, adding the same constant to all options should not change the choice outcome. In this paper, we show that the normalisation of SB is *not* choice consistent in this sense and propose an alternative normalisation to determine the dissimilarity parameter λ.

Based on a choice consistent normalisation, we then introduce the dissimilarity parameter λ to the Poisson count framework using aggregate location choice data. From this, with panel data, the dissimilarity parameter can be computed from the group effects of the Poisson count regression without specific knowledge about the number of times the outside option has been chosen. Intuitively, the group effects of a Poisson count regression with panel data absorb the discrepancy between the observed number of location choices and the expected number from a basic Poisson process, and hence provide clues about the relative importance of the (unobserved) outside option.

For several reasons, the value of the dissimilarity parameter can be important. Firstly, for a given sample, it indicates how far empirical location choices reflect a positive or zero sum world. Secondly, the dissimilarity parameter permits us to paint a more nuanced picture when calculating the resulting elasticities by taking into account such things as (i.) differences of elasticities across locations (ii.) the magnitude of spillover effects when the economic or political conditions change elsewhere, or (iii.) the degree of similarity of locations competing to attract a firm.

There are many applications where the differences between the various location choice models could matter. We will apply these models to an example where location choices are revealed in CBA deals within a sample of 25 EU countries. The reason is that this data is comprehensively available and the location choices of MNEs lend themselves to the introduction of an outside option that could represent such contingencies as exporting to a given market instead.

The paper is organised as follows. The first part reviews the contributions of Guimarães *et al.* (2003) and SB that are relevant for the current context. In particular, to prepare the ground, section 2 discusses the similarities and differences between the basic conditional logit model and Poisson count regression. Section 3 provides a synoptic overview of the nested logit model with a fixed outside option suggested by SB. Section 4 discusses the role of choice consistent normalisations and, based on this, section 5 shows how the dissimilarity parameter λ can be computed within a Poisson count framework. Section 6 discusses the empirical application. Section 7 concludes.

2 Basic Location Choice Models

Consider the case where a firm wants to place some economic activities in a given location. Let the firms that are observed to undertake a location choice be indexed with $i = 1, \ldots, N$. The domicile of the investing firms, or "source" of the investment, is denoted with $s = 1, \ldots, S$ whilst the choice set includes potential target locations, or "hosts" of the investment, indexed with $h = 1, \ldots, H$. With profit maximising firms, the observed location choice—which is henceforth denoted by $l_{i,sh}$—reveals that the location h with profit opportunity $E[\Pi_{i,sh}]$ is expected to outperform all competing alternatives h' that could in principle have been chosen instead. Hence,

$$l_{i,sh} = \begin{cases} 1 & E[\Pi_{i,sh}] > E[\Pi_{i,sh'}] \quad \forall \quad h \neq h' \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The conditional logit model employs location choices such as (1) as the dependent variable. Thereby, a set of choice-specific variables x_{sh} is thought to impact upon profit expectations $E[\Pi_{i,sh}]$ through the linear function

$$E[\Pi_{i,sh}] = \delta_s + x'_{sh}\beta + \epsilon_{i,sh}, \tag{2}$$

where δ_s is a group effect absorbing source-specific factors. Furthermore, β denotes the coefficients to be estimated. The component $\epsilon_{i,sh}$ accounts for stochastic factors that impact upon the location choice. To reflect that (1) uncovers the location with the highest expected profit opportunity, $\epsilon_{i,sh}$ is usually assumed to follow a Gumbel, or type 1 extreme value distribution (McFadden, 1974) where the location and scale parameter have been normalised to, respectively, 0 and 1.[1] The probability that a firm of s chooses h is then given by the corresponding ratio between their expected number $E[n_{sh}]$ and the total number $E[\overline{N}] = \sum_{s=1}^{S}\sum_{h=1}^{H} E[n_{sh}]$ of location choices within the sample, that is

$$P_{sh} = \frac{\exp\left(x'_{sh}\beta\right)}{\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)} = \frac{E[n^{cl}_{sh}]}{E[\overline{N}]}. \tag{3}$$

Owing to the logistic structure of (3), components such as δ_s that are fixed across the different options drop out. Taking the joint distribution over all observed firms i, sources s, and host locations h yields the log likelihood function

$$\ln L^{cl}(\beta) = \sum_{s=1}^{S}\sum_{h=1}^{H} n_{sh}P_{sh} = \sum_{s=1}^{S}\left\{\sum_{h=1}^{H} n_{sh}x'_{sh}\beta - \sum_{h=1}^{H}\left[n_{sh}\ln\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]\right\} \tag{4}$$

from which the coefficients β can be estimated. Guimarães *et al.* (2003) show that a direct regression onto the count variable n_{sh} aggregating the number of location choices across sources and hosts provides an often more convenient way to estimate the coefficients β. This becomes clear when multiplying (3) with the denominator, which yields a conventional (panel) count regression with exponential mean transformation of

$$E[\widetilde{n}^{pc}_{sh}] = \exp\left(\delta_s + x'_{sh}\beta\right) = \alpha_s E[n^{pc}_{sh}], \tag{5}$$

where $\alpha_s = \ln(\delta_s)$ and $E[n^{pc}_{sh}] = \exp(x'_{sh}\beta)$. Here, the group effect δ_s has been retained, but in principle the transformation between conditional logit and Poisson count regression works regardless of whether or not δ_s is included in (2). Assuming that the exponential mean parameter $E[\widetilde{n}^{pc}_{sh}]$ is Poisson distributed with probability density

[1] The Gumbel distribution, which is sometimes also referred to as type 1 extreme value distribution, has a cumulative distribution function of $F(\ell_s, \varsigma_s) = \exp(-\exp(-(\epsilon_{i,sh} - \ell_s)/\varsigma_s))$ where ℓ_s is the the so-called location and $\varsigma_s > 0$ the scale parameter. For more details, see e.g. Ben-Akiva and Lerman (1985) or Hensher *et al.* (2005, ch.13 and 14).

$$P[\widetilde{n}^{pc}_{sh}] = \frac{\exp(-E[\widetilde{n}^{pc}_{sh}])(E[\widetilde{n}^{pc}_{sh}])^{\widetilde{n}_{sh}}}{\widetilde{n}^{pc}_{sh}!} \tag{6}$$

yields coefficient estimates for β that are identical to those of the conditional logit model. To see why, contemplate the log likelihood contribution of s, that is

$$\ln L^{pc}_{s}(\alpha_s, \beta) = -\alpha_s \sum_{h=1}^{H} \exp(x^{'}_{sh}\beta) + \ln \alpha_s \sum_{h=1}^{H} n_{sh} + \sum_{h=1}^{H} n_{sh} x^{'}_{sh}\beta - \sum_{h=1}^{H} \ln n_{sh}! \tag{7}$$

Setting the first derivative of this with respect to α_s equal to 0, and solving for α_s yields[2] the maximum likelihood estimator of

$$\alpha_s = \frac{\sum_{h=1}^{H} n_{sh}}{\sum_{h=1}^{H} \exp\left(x^{'}_{sh}\beta\right)} = \frac{\overline{n}_s}{E[\overline{n}_s]}. \tag{8}$$

Hence, for source s, α_s absorbs the discrepancy between the observed number of aggregated location choices $\overline{n}_s = \sum_{h=1}^{H} n_{sh}$ and the corresponding expected number $E[\overline{n}_s] = \sum_{h=1}^{H} \exp(x_{sh}\beta)$ from the basic Poisson distribution. Thereby, $0 < \alpha_s < 1$ and $1 < \alpha_s$ represent cases where, respectively, the number of observed location choices are "underreported" and "overreported" relative to the basic Poisson count distribution. Substituting (8) back into (7) and summing over all H hosts yields the concentrated log-likelihood function (that no longer depends on α_s) given by

$$\ln L^{pc}(\beta) = \sum_{s=1}^{S} \left\{ \sum_{h=1}^{H} n_{sh} x^{'}_{sh}\beta - \sum_{h=1}^{H} \left[n_{sh} \ln \sum_{h=1}^{H} \exp\left(x^{'}_{sh}\beta\right)\right]\right\} + \text{constant}, \tag{9}$$

which differs from (4) only as regards a constant with respect to β. Hence, the corresponding estimates are identical.

The variables x_{sh} enter into the conditional logit model and the Poisson count regression in a non-linear manner. Hence, the coefficients β do not reflect a marginal effect (Hensher *et al.*, 2005, pp.383ff.). This warrants the calculation of the elasticity η to reflect the percentage change of the expected number $E[n_{sh}]$ of location choices between s and h in response to a percentage change of a given variable $x_{sh,k}$. Though the same coefficients arise from the conditional logit model and the Poisson count regression, SB have recently drawn attention to the differences in the resulting elasticities. Specifically, when expressing the variables in logarithms, the own-elasticity of the Poisson count regression equals

$$\eta^{pc}_{k} = \frac{\partial E[\widetilde{n}^{pc}_{sh}]}{\partial x_{sh,k}} \frac{x_{sh,k}}{E[\widetilde{n}^{pc}_{sh}]} = \beta_k \tag{10}$$

where β_k denotes the coefficient pertaining to $x_{sh,k}$ and $E[\widetilde{n}^{pc}_{sh}]$ is given by (5). Recall that in the conditional logit model, individual location choices $l_{i,sh}$ are the dependent variable. This implies that the total number of location choices $\overline{N}$ is fixed—in the sense of not depending on x_{sh}—and the expected number of location choices between s and h is the probability weighted expression $E[n^{cl}_{sh}] = \overline{N}P_{sh}$. Inserting (3) for P_{sh} yields an elasticity of

$$\eta^{cl}_{sh,k} = \frac{\partial E[n^{cl}_{sh}]}{\partial x_{sh,k}} \frac{x_{sh,k}}{E[n_{sh}]} = (1 - P_{sh})\beta_k. \tag{11}$$

Owing to the properties of the probability $P_{sh} \in [0, 1]$, this is not larger than (10).

The different elasticities are a result of polar assumptions as regards the degree of similarity between different locations. In the Poisson count regression, the hosts are thought to

[2] A textbook discussion and derivation of the Poisson count regression with panel data can be found e.g. in Winkelmann (2008, ch.7.2) or Cameron and Trivedi (1998, ch.9).

represent alternatives that are dissimilar, or segmented, in the sense that a changing value of the variable $x_{sh,k}$ affects the conditions in h, but not elsewhere in h'. The absence of such spillovers manifests, statistically, in a cross-elasticity that is equal to zero by definition.[3] Borrowing the words of SB, this implies that the Poisson count regression reflects a "positive sum world" where e.g. improving conditions in h can result in an expansion of the total number of location choices $N = \sum_s \sum_h n_{sh}$. Conversely, the conditional logit model is a "zero sum world" where the total number of location choices is fixed at $\overline{N}$. When more firms choose h, this comes entirely at the expense of competing alternatives within the set H of locations that are deemed to be similar.[4]

It is not always appropriate to restrict the alternative of locating in h to locating elsewhere as in the conditional logit model, or ignoring the role of alternatives altogether. Consider for example the case of FDI. As depicted in a schematic manner in figure 1, a firm can either decide to go multinational indexed with $m = \emptyset$ and, contingent on this, choose a host h to locate a subsidiary plant. Alternatively, as depicted by the branch on the left, a firm can also invest in a generic outside option $m = o$, which comprises only one elemental choice, labelled with $h = 0$, encompassing contingencies such as remaining inactive, exporting, installing capacity in the home country, etc.

FIGURE 1 HERE

Accounting explicitly for an outside option has other benefits. In particular, SB show that the extension of the conditional logit model with a fixed outside option yields a version of the nested logit model that covers the intermediate cases between the zero and positive-sum world discussed above. The next section endeavours to develop this model within the current context.

3 Location Choice in a Nested Logit Model

Following SB, and as depicted in figure 1, consider a scenario where the outside option $h = 0$ is fixed in the sense of not depending on the choice-specific variables x_{sh}. Hence, the corresponding expected profit is given by

$$E[\Pi_{i,s0}] = \delta_s + \epsilon_{i,s0}. \tag{12}$$

Similar to (1), a firm is again assumed to pursue the outside option $h = 0$ if the resulting profit is expected to outperform the alternative of placing economic activities in a given location $h > 0$, with the corresponding decision being denoted by

$$m_{i,sh} = \begin{cases} 0 & E[\Pi_{i,s0}] > E[\Pi_{i,sh}] \quad \forall \quad h > 0 \\ 1 & \text{otherwise.} \end{cases} \tag{13}$$

The nested logit model entertains the idea that the elemental options can be summarised into different groups consisting, here, of the outside option $h = 0$ and the location choice $h > 0$, whereby the elemental options are supposed to be more similar within than between these different "nests". To embed the approach of SB within the present context, the following provides a synoptic derivation of the choice probabilities of a nested logit model reflecting the structure of figure 1.[5]

[3] Deriving (5) with respect to x'_{sh} yields $\zeta^{pc}_{sh,k} = \frac{\partial E[\tilde{n}^{pc}_{sh}]}{\partial x_{sh'k}} \frac{x_{sh'k}}{E[\tilde{n}^{pc}_{sh}]} = 0$.

[4] Statistically, this manifests in a cross-elasticity of $\zeta^{cl}_{sh,k} = -P_{sh}\beta_k$. Summing over all options H yields $(1 - P_{sh})\beta_k - (H - 1)P_{sh}\beta_k = (1 - HP_{sh})\beta_k = 0$ and uncovers the zero-sum property in the sense that any "gain" of choices h is exactly offset by the "losses" in the other options.

[5] For a textbook discussion of the nested logit model see Hensher *et al.* (2005, ch.13-14).

Consider first the basic location choice when a firm i does not want to pursue the outside option (where $m_{i,sh} = 1$). Recall that the expected profits of (2) guiding the location choice have a stochastic component to account for unobservable factors. However, in contrast to the conditional logit model, $\epsilon_{i,sh|ø}$ conditions here on the firm not pursuing the outside option before making elemental decisions about the host $h > 0$. The stochastic component $\epsilon_{i,sh|ø}$ is again assumed to be Gumbel-distributed with a location parameter normalised to 0, but a variable scale parameter $\varsigma_s^ø > 0$ reflecting the similarity of the elemental options within the group. Now, the conditional probability that a location $h > 0$ is chosen is given by

$$P_{sh|ø} = \frac{\exp\left(x_{sh}^{'}\beta\varsigma_s^ø\right)}{\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\varsigma_s^ø\right)}. \tag{14}$$

This differs from (3) only with respect to the scale parameter $\varsigma_s^ø$, which was normalised to 1 in the conditional logit model (see section 2). Turning to the group stage, the probability $P_{øs}$ that the outside option is *not* pursued depends again on a Gumbel distribution with scale parameter $\lambda_s^ø$. From this, the probability $P_{øs}$ is given by

$$P_{øs} = \frac{\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\varsigma_s^ø\right)\right]^{\frac{\lambda_s^ø}{\varsigma_s^ø}}}{\left[\exp\left(\delta_s\varsigma_s^o\right)\right]^{\lambda_s^o} + \left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\varsigma_s^ø\right)\right]^{\frac{\lambda_s^ø}{\varsigma_s^ø}}}, \tag{15}$$

where the first term of the denominator accounts for the probability contribution of the outside option (see below). The extent to which the options in the location choice group differ manifests itself in the dissimilarity parameter[6] $(\lambda_s^ø/\varsigma_s^ø) \in [0,1]$ which weights the probability contribution of $E[N^ø] = \sum_{s=1}^{S}\sum_{h=1}^{H}\exp(x_{sh}^{'}\beta\varsigma_s^ø)$ and is called "inclusive value" since it connects the elemental location choices of (14) with the group choice stage of (15). It can be shown that

$$(\lambda_s^ø/\varsigma_s^ø) = \sqrt{1-\rho_s^ø} \tag{16}$$

where $\rho_s^ø \in [0,1]$ is the correlation between the stochastic components $\epsilon_{i,sh|ø}$ pertaining to the profits from investing in different locations.[7] Jointly, (14) and (15) define the unconditional probability of locating economic activities in $h > 0$, that is

$$P_{sh} = P_{øs} \times P_{sh|ø} = \frac{\exp\left(x_{sh}^{'}\beta\varsigma_s^ø\right)\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\varsigma_s^ø\right)\right]^{\left(\frac{\lambda_s^ø}{\varsigma_s^ø}-1\right)}}{\left[\exp\left(\delta_s\varsigma_s^o\right)\right]^{\lambda_s^o} + \left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\varsigma_s^ø\right)\right]^{\frac{\lambda_s^ø}{\varsigma_s^ø}}}. \tag{17}$$

Consider now the fixed outside option o that represents a degenerated nest in the sense of offering only one basic "choice" with $h = 0$. As discussed in Hunt (2000), in this case, the distinction between unconditional and conditional probabilities is irrelevant, as $P_{s0|o} = 1$ and $P_{s0} = P_{os} \times P_{s0|o}$. Since $P_{os} = 1 - P_{øs}$ for the present binary choice with $P_{øs}$ defined in (15), the probability of choosing the outside option is given by

$$P_{os} = P_{s0} = \frac{\exp\left(\delta_s\varsigma_s^o\right)^{\lambda_s^o}}{\left[\exp\left(\delta_s\varsigma_s^o\right)\right]^{\lambda_s^o} + \left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\varsigma_s^ø\right)\right]^{\frac{\lambda_s^ø}{\varsigma_s^ø}}}. \tag{18}$$

[6]Sometimes this is also referred to as inclusive value parameter or log-sum coefficient.
[7]A derivation of this result can be found in Hunt (2000, p.98).

The coefficients β can be estimated by means of maximum likelihood from the joint probabilities (17) and (18) across observed firms i. However, empirically, only the correlation $\rho_s^{\o}$ can be estimated from the data, but not the scale parameters $\lambda_s^{\o}$ and $\varsigma_s^{\o}$ (Hunt, 2000; Ben-Aktiva and Lerman, 1985; Hensher *et al.*, 2005, ch.13). In essence, this represents an over-identification problem that necessitates some normalisation. Usually, this involves setting some scale parameters at the group or basic choice level to 1 or 0. SB (p.217) set $\varsigma_s^{\o} = 1$, $\varsigma_s^{o} = 1$, and $\lambda_s^{o} = 1$ wherefore the probability of investing in the fixed outside option $h = 0$ according to (17) or of choosing location $h > 0$ according to (18) becomes

$$P_{sh} = \frac{\exp\left(x_{sh}^{'}\beta\right)\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]^{(\lambda_s^{\o}-1)}}{\exp\left(\delta_s\right)+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]^{\lambda_s^{\o}}} = \frac{\exp(x_{sh}^{'}\beta)(E[N^{\o}])^{(\lambda_s^{\o}-1)}}{\exp(\delta_s)+(E[N^{\o}])^{\lambda_s^{\o}}} \quad (19)$$

$$P_{0s} = \frac{\exp\left(\delta_s\right)}{\exp\left(\delta_s\right)+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]^{\lambda_s^{\o}}} = \frac{\exp(\delta_s)}{\exp(\delta_s)+(E[N^{\o}])^{\lambda_s^{\o}}} \quad (20)$$

where $E[N^{\o}] = \sum_{s=1}^{S}\sum_{h=1}^{H}\exp(x_{sh}^{'}\beta)$ is the inclusive value when imposing the above mentioned normalisation. Then, for firms domiciled in s, denoting the variable counting the number of times that the outside option has been chosen with n_{os} and the number of times where this is not the case with $n_{\o s}$, the concentrated log-likelihood function of the nested logit model with a fixed outside option, as derived in Appendix A.3 of SB, is given by

$$\ln L^{nl}(\beta) = \sum_{s=1}^{S}\left\{\sum_{h=1}^{H} n_{sh}x_{sh}^{'}\beta - \sum_{h=1}^{H}\left[n_{sh}\ln\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]\right\} + \text{constant}. \quad (21)$$

Again, this differs from the Poisson count regression and the conditional logit model only up to a constant and hence yields identical estimates for the coefficients β. Furthermore, according to SB (2011, p.217), the elasticity of the nested logit model is given by

$$\eta_{sh,k} = \frac{\partial E[n_{sh}]}{\partial x_{sh,k}} = [1 - P_{sh|\o}(1-\lambda_s^{\o}P_{s0})]\beta_k, \quad (22)$$

which coincides with the basic conditional logit model when $\lambda_s^{\o} = 0$ and with the basic Poisson count regression when $\lambda_s^{\o} = 1$ and $P_{s0} = 1$.

4 A Choice Consistent Normalisation

Though adding a fixed outside option leads to a location choice model encompassing the Poisson count and the conditional logit framework, the nested logit approach of section 3 suffers from several drawbacks.

Firstly, the normalisation of the scale parameters λ and ς is a critical step in the sense of leading to different versions of the nested logit model with different results and elasticities (Hunt, 2000; Hensher and Greene, 2005). Arguably, this aspect is often neglected in applied work (Louviere *et al.*, 2000; Hensher *et al.*, 2005, p.538). To avoid ambiguities, Koppelman and Wen (1998) suggest that the normalisation should be consistent with some plausible principles of choice theory. For example, since adding the same constant Δ to all profits of (2) and the outside option (12) would not change the ranking of the elemental options, a theoretically consistent nested logit model should be invariant to such a transformation. However, appendix A shows that the normalisation underlying (19) and (20) does not fulfill this property.

Secondly, the scale parameter λ_s does not appear in the concentrated log likelihood function (21) and, under the normalisation imposed in section 3, the maximum likelihood estimate for δ_s and λ_s appear in the same first order condition

$$\exp(\delta_s) = \frac{n_{os}}{n_{\o s}} \left[\sum_{h=1}^{H} \exp\left(x_{sh}^{'}\beta\right) \right]^{\lambda_s} \tag{23}$$

and hence cannot be separately identified (SB, 2011, p.217).

We address these caveats by considering an alternative, choice consistent normalisation:

Proposition 1 *Setting $\lambda_s^o = \lambda_s^{\o} = \lambda_s$, $\varsigma_s^{\o} = 1$, and $\varsigma_s^o = 0$ represents a normalisation that is consistent with choice theory in the sense that adding a constant Δ to the profits (2) and (12) that guide the elemental choices does not change the choice outcome.*

PROOF: Appendix A.

With the normalisation of proposition 1, the probability of opting for the outside option according to (17), or of choosing location $h > 0$ according to (18) becomes

$$P_{sh} = \frac{\exp\left(x_{sh}^{'}\beta\right)\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]^{(\lambda_s-1)}}{1+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]^{\lambda_s}} = \frac{(E[N^{\o}])^{(\lambda_s-1)}}{1+(E[N^f])^{\lambda_s}}\exp\left(x_{sh}^{'}\beta\right) \tag{24}$$

$$P_{s0} = \frac{1}{1+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]^{\lambda_s}} = \frac{1}{1+(E[N^{\o}])^{\lambda_s}}, \tag{25}$$

Proposition 1 implies that $\exp(\delta_s\varsigma_s^o)^{\lambda_s^o} = \exp(0)^{\lambda_s} = 1$ and, hence, normalises to 1 the contribution of the outside option.[8] Intuitively, this is not problematic since the introduction of the scale parameter λ_s to the expected number of location choices $E[N^{\o}]^{\lambda_s}$, which appears in the denominator of (24) and (25), already weights the relative importance between the outside option and the location choices. Also, the normalisation of proposition 1 leaves the likelihood function (21) intact since the parameter δ_s does not appear in it.

When $\varsigma_s^{\o} = 1$, according to (16), the scale parameter maps into the correlation between the stochastic component $\epsilon_{i,sh}$ of the elemental options, that is $\lambda_s = \sqrt{1-\rho_s}$, and hence reflects directly the degree of dissimilarity between the different locations. Recall that $\rho_s = 0$ means that stochastic events are entirely uncorrelated and the different locations are completely segmented, which is consistent with the basic Poisson count regression with dissimilarity parameter $\lambda_s = 1$. Conversely, a dissimilarity parameter of $\lambda_s = 0$, which implies that $\rho_s = 1$, means that the locations are perfectly integrated, which is consistent with the conditional logit model.

Though with proposition 1, which implies that $\exp(\delta_s\varsigma_s^o)^{\lambda_s^o} = \exp(0)^{\lambda_s} = 1$, the dissimilarity parameter λ_s could be estimated from (23), this would still necessitate information about the number of times n_{os} that the outside option has been chosen. Considering again the example of FDI, this is not straightforward since contingencies such as abandoning an investment project or remaining inactive are hard to observe. However, the next section suggests that aggregating location choices provides a possible remedy, when the outside option is (partly) unobservable.

[8] A similar normalisation arises with the Logit model for binary choices where the probability of choosing alternative 1 and 2 are, respectively, $\exp(x^{'}\beta)/(1+\exp(x^{'}\beta))$ and $1/(1+\exp(x^{'}\beta))$ meaning that the weight of one alternative has also been normalised to 1.

5 Introducing λ_s into the Poisson Count Framework

Similar to the basic models of section 2, the following endeavours to establish the link between the nested logit model, where individual location choices are the dependent variable, and a corresponding Poisson count regression, where the aggregate number of such choices is the dependent variable, when adding a fixed outside option to the choice set. Following the steps of section 2 to obtain a model with count variable n_{sh} as dependent variable, multiplying the nested logit probability P_{sh} of (24) with the expected total number of firms $E[N] = 1 + (E[N^{ø}])^{\lambda_s}$ of the denominator yields

$$E[\widetilde{n}_{sh}^{pcu}] = P_{sh}E[N] = \underbrace{(E[N^{ø}])^{(\lambda_s - 1)}}_{=\alpha_s} \exp\left(x_{sh}^{'}\beta\right) = \alpha_s E[n_{sh}^{pc}], \tag{26}$$

which has a similar structure to (5), but the group effect that can be estimated from (8) has been parameterised by $\alpha_s = E[N^{ø}]^{(\lambda_s-1)}$. When $\lambda_s = 1$, we have $E[N^{ø}]^{(\lambda_s-1)} = E[N^{ø}]^0 = 1$ and the basic Poisson count regression with completely dissimilar elemental options arises. Then, the inclusion of an outside option is irrelevant, which manifests itself in the fact that $E[\widetilde{n}_{sh}^{pcu}]$ coincides with $E[n_{sh}^{pc}]$. Conversely, when $\lambda_s < 1$, the elemental options are to some degree similar and, the more this is the case, the more the expected number $E[\widetilde{n}_{sh}^{pcu}]$ of location choices differ from $E[n_{sh}]$ of a basic count process.

Solving $\alpha_s = E[N^{ø}]^{(\lambda_s-1)}$ for the dissimilarity parameter yields

$$\widehat{\lambda}_s = \frac{\ln(E[N^{ø}]\alpha_s)}{\ln(E[N^{ø}])}. \tag{27}$$

Recall from the discussion in section 3 that λ_s can adopt values between 0 and 1. Then, $\lambda_s = 1$, which is consistent with the basic Poisson count regression, arises when $\alpha_s = 1$. Conversely, $\lambda_s = 0$, which is consistent with the basic conditional logit model, requires that $\alpha_s = 1/E[N^{ø}]$ which is close to 0 when a large number $E[N^{ø}]$ of location choices are expected in the data. According to (8), when $0 < \alpha_s \leq 1$, the count data exhibit underreporting in the sense that the observed number of location choices $\widetilde{n}_{sh}^{pcu}$ is lower than would be expected from a basic Poisson count process with n_{sh}^{pc}. Unless $\alpha_s = 1$, some firms will indeed end up choosing the (unobserved) outside option, which reduces the number of location choices actually observed. Hence, with aggregate counts, a high degree of underreporting can be interpreted as evidence of a higher importance of the outside option. Note that the establishment of this nexus between the nested logit model and the Poisson count regression of (5) necessitates panel data to obtain the group effect (8) and compute the dissimilarity parameter λ_s in (27).

Finally, as derived in appendix B, the elasticity of $\widetilde{n}_{sh}^{pcu}$ with respect to changes in $x_{sh,k}$ with coefficient β_k now equals

$$\eta_{sh,k}^{pcu} = [1 - P_{sh|ø}(1 - \lambda_s)]\beta_k. \tag{28}$$

This is again entirely consistent with the framework above in the sense that $\lambda_s = 1$ returns the elasticity of the basic Poisson count regression given by (10) and $\lambda_s = 0$ is the elasticity of the conditional logit model as given by (11). However, empirically, the dissimilarity parameter can adopt any value between these polar cases. Also, evaluating (28) yields $\eta_{sh,k}^{pcu} = \lambda_s\beta_k + (1-\lambda_s)(1-P_{sh|ø})\beta_k = \lambda_s\eta_{sh,k}^{pc} + (1-\lambda_s)\eta_{sh,k}^{cl}$, meaning that the elasticity of the Poisson count regression with a fixed outside option is a with λ_s weighted linear average between the basic Poisson count regression and the conditional logit model. This reflects a similar condition for the nested logit model that features in SB (2011,pp.217ff.). In sum, the introduction of an outside option leads to a more nuanced picture when calculating the resulting elasticities since the value of (28) depends on (i.) the coefficient β_k which determines the upper bound of the elasticity, (ii.) the probability $P_{sh|ø}$ that a host h can attract a firm from elsewhere when changing the value of $x_{sh,k}$, and (iii.) the extent to which the locations are similar and hence compete to attract firms.

6 Empirical Application: Cross-Border Acquisitions

This section endeavours to illustrate the method from section 5 to calculate the dissimilarity parameter λ_s from the data with an application drawing on the location choices revealed when firms acquire a subsidiary plant abroad. Such CBAs are comprehensively recorded in the SDC Platinum database of Thomson Reuters and have been used elsewhere to study the determinants of FDI within the Poisson count framework (Kessing *et al.*, 2007; Herger *et al.*, 2008; Hijzen *et al.*, 2008; Coerdacier *et al.*, 2009) and the conditional logit framework (Herger *et al.*, 2011). To focus on a group, or nest, with relatively similar locations, the CBA deals between 25 EU countries during the 2005 to 2009 period are used. In total, the sample contains 8,302 deals with the top panel of Table 1 recording the count (number of CBAs) between the source, reported in rows, and host, reported in columns.

TABLE 1 HERE

Following the literature on the determinants of FDI, the profit function (2) guiding the location choice is fitted to a gravity equation with the list of dependent variables x'_{sh} containing real GDP, the WAGE level, the distance (DIST) between the capital cities, a dummy variable for countries sharing a common BORDER, a common language dummy variable (LANG), a measure for investment freedom (INVFR), an index on trade freedom (TRADEFR), a corruption index (CORRUPT), and a measure for the effective average taxes levied on corporations (TAX). Table 3 of appendix C contains a detailed description and the sources of the variables. Except for the dummy variables, the regressors have been transformed into logarithms and averaged over the 5 years under consideration. The resulting coefficients, calculated from a panel Poisson count regression with group effects α_s and with the 600 (25 source $\times$ 24 hosts) observations of n_{sh} in the top panel of Table 1 as dependent variable, are given by

$$\begin{aligned}\#CBA_{sh} \quad &= \quad \alpha_s * \exp\Bigg(\underset{\substack{(0.02)\\ [0.02]}}{0.74}GDP - \underset{\substack{(0.06)\\ [0.06]}}{1.04}WAGE - \underset{\substack{(0.02)\\ [0.02]}}{0.58}DIST + \underset{\substack{(0.03)\\ [0.03]}}{0.51}BORDER + \underset{\substack{(0.04)\\ [0.04]}}{0.67}LANG \\ &\quad + \underset{\substack{(0.10)\\ [0.10]}}{0.60}INVFR - \underset{\substack{(0.87)\\ [0.88]}}{3.89}TRADEFR - \underset{\substack{(0.10)\\ [0.10]}}{1.37}CORRUPT - \underset{\substack{(0.10)\\ [0.10]}}{0.40}TAX \Bigg), \end{aligned} \tag{29}$$

whereby standard deviations are reported below the coefficients in (round) parentheses. Note that all coefficients are significant and shape up to the economic priors with more CBAs occurring with economically large hosts that have low wages, are geographically close to the source, have a common border or language, offer a high degree of investment freedom, are difficult to access by trade, and have low levels of corruption and corporate taxes. Recalling the discussion in section 2, a conditional logit model employing location choices $l_{i,sh}$ revealed in individual CBA deals yields the same coefficient estimates. However, in practice, it is more burdensome to handle such a location choice model since it involves 199,248 observations (24 possible choices $\times$ 8,302 deals) rather than the 600 when using aggregated count data. The standard deviations of the conditional logit model are reported in [square] brackets in (29) and barely differ, here, from those of the Poisson count process and do not overturn the significance of any of the coefficient estimates. Conversely, since (7) contains a constant that does not appear in (4), the value of the log likelihood function of the Poisson count regression, given by -2,139, is much less negative than the value of the conditional logit model given by -21,741.

The estimated group effects $\widehat{\alpha}_s$ are reported in Table 2. Note that the values of $\widehat{\alpha}_s$ are all below $1/3$ which suggests that the number of observed CBA deals is substantially lower than would be expected from a basic Poisson count regression without adjusting for the group effect of (8). The relatively low values of α_s suggest that, within the current sample of EU countries, firms often seem to favour the generic outside option.

TABLE 2 HERE

From the group effects $\widehat{\alpha}_s$ of Table 2, the dissimilarity parameter λ_s can be computed from (27), whereby, from the estimates of (29), $E[N^{\o}] = \sum_s \sum_h E[n_{sh}] = \sum_s \sum_h \exp(x'_{sh}\beta) = 132,848$. In general, for the current example, the values of the dissimilarity parameters are closer to the bound with $\lambda_s = 1$ that reflects the "positive sum world" of the Poisson count regression. This would suggest that additional CBAs with a given host country do often not come at the expense of alternative locations. However, some differences arise between the countries. What may be worth noting is that the dissimilarity parameter tends to be lower for the Eastern European countries that have only recently joined the EU. Perhaps, these countries have attracted a relatively large proportion of firms trying to outsource production stages to low wage countries, which could be a type of investment that is subject to more rivalry between locations and is hence closer to the "zero sum world" of the conditional logit model.

The value of the dissimilarity parameter also has implications for the calculation of elasticities. To see this, recall, from (10), that the coefficient estimates β_k of (29) represent the upper bound of the absolute value of an elasticity η_k^{pc} reflecting the positive-sum world inherent in the Poisson count regression. Conversely, in the zero-sum world of the conditional logit model, which determines the lower bound $\eta_{sh,k}^{cl} = (1 - P_{sh})\beta_k$ of (11), the elasticity depends also on the probability P_{sh} that a firm domiciled in s invests in h. When the coefficient β_k is low, such as the impact of taxation in (29) with an inelastic value of -0.40, or the probability P_{sh} is modest, which is the case here for many country pairs as reported in the lower panel of Table 1, the range between these bounds is small. It may be intuitive that distinguishing between the zero and positive-sum world might be irrelevant when a variable is, statistically and economically, insignificant or a location is unlikely to attract a firm anyway, implying that changes of the local conditions have negligible spillover effects on to other locations. Conversely, when a variable has a potentially large impact, such as corruption which enters into (29) with a coefficient of -1.39, and the probability of investment between s and h is large, such as from Ireland into the UK where $P_{sh} = 0.49$, the lower bound would e.g. be $\eta_{sh,k}^{cl} = (1 - 0.49)(-1.39) = -0.71$ meaning about half the maximal value. Then, it might be important to pin down the position of the intermediate elasticity of (28). Considering again the case of corruption and using the range arising for a CBA from Ireland into the UK, with the relevant dissimilarity parameter being $\lambda_s = 0.74$ according to Table 2, applying equation (28) yields an elasticity of $\eta_{sh,k}^{pcu} = 0.74 \times (-1.37) + (1 - 0.74) \times (-0.71) = -1.20$. Hence, in this case, the elasticity would be closer to the upper bound.

Though the example presented in this section yields some intuitive results, our aim is merely to illustrate how to calculate the dissimilarity parameter λ_s from the data with the method from section 5. Of course, variables other than those in (29) have appeared in the literature and considering them might affect the results. In any case, considering different location choice models and including the contingency of a fixed outside option leads to a more versatile picture e.g. in terms of an elasticity whose value is not uniform across locations, but depends also on the conditions elsewhere or on the degree of similarity between competing locations.

7 Summary and Conclusion

Econometric models employing discrete location choices as the dependent variable have become a popular framework for uncovering how economic, political, and other determinants affect the geographical distribution of economic activities. Specifically, the econometric analysis of location choices has taken the form of a conditional logit model, which connects the determinants with individual decisions to locate economic activities in a given place, or a Poisson regression aggregating such location choices into a count variable. Though they yield identical coefficient estimates, the conditional logit model and the Poisson count regression reflect polar cases when it comes to the interpretation of the results. Above all, in the Poisson count regression, the locations are thought to be completely dissimilar options whilst in the conditional logit model they are completely similar. Previous work by Schmidheiny and Brülhart (2011) suggests that these polar cases can be reconciled by adding a fixed outside option and transforming the conditional into a nested logit model. This gives rise to a dissimilarity parameter λ_s whose value covers the continuum between the Poisson count regression and the conditional logit model. However, outside options that cannot be observed as well as data handling issues when there are a large number observations and locations, can inhibit the empirical estimation of the dissimilarity parameter. For the case of panel data, this paper has shown that the outside option can also be introduced into the Poisson count framework, where a group effect accounts for the underrecording of location choices and provides a way to uncover the value of the dissimilarity parameter. The main advantages of using such a Poisson count framework are that (1.) location choices taken for example by firms are often easier to observe than, say, the value they invest and (2.) aggregating the location choices into a count variable can lead to a dramatic reduction in the number of observations required for estimation.

References

Ben-Aktiva, M., and S.R. Lerman (1985), Discrete Choice Analysis: Theory and Application to Travel Demand, Cambridge: MIT Press.

Buettner, T. and M. Ruf (2007), Tax incentives and the location of FDI: evidence from a panel of German multinationals, International Tax and Public Finance 14: 151-164.

Cameron, C.A. and P.K. Trivedi (1998), Regression analysis of count data, Cambridge, Cambridge University Press.

Coerdacier, N., R. De Santis, and A. Aviat (2009), 'Cross-Border Mergers and Acquisitions and European Integration', Economic Policy, 55-106.

Crozet, M., T. Mayer, and J.L. Mucceilli (2004), How do firms agglomerate? A study of FDI in France, Regional Science and Urban Economics 34: 27-54.

Devereux, M.P., C. Elschner, D. Enders, and C. Spengel (2009), 'Effective Tax Levels using the Devereux/Griffith Methodology', Mannheim and Oxford: Centre for European Economic Research.

Devereux, M. and R. Griffith (1998), Taxes and the location of production: evidence from a panel of US multinationals, Journal of Public Economics 68: 335-367.

Devereux, M., R. Griffith and H. Simpson (2007), Firm location decisions, regional grants and agglomeration externalities, Journal of Public Economics 91: 413-435.

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

Hensher, D.A., J.M.Rose, and W.H.Greene (2005), Applied Choice Analysis, Cambridge: Cambridge University Press.

Hunt, G.L. (2002), Alternative nested logit model structures and the special case of partial degeneracy, Journal of Regional Science 40: 89-113.

Herger, N., C. Kotsogiannis and S. McCorriston (2008), Cross-Border Acquisitions in the Global Food Sector, European Review of Agricultural Economics 35: 563-587.

Herger, N., C. Kotsogiannis and S. McCorriston (2008), International Taxation and FDI Strategies: Evidence from US Cross-Border Acquisitions, University of Exeter Economics Department Discussion Paper 11/09.

Hijzen, A., H. Görg and M. Manchin (2008) Cross-border mergers & acquisitions and the role of trade costs, European Economic Review 52: 849-866.

Kessing, S., K.A. Konrad and C. Kotsogiannis (2007), Foreign direct investment and the dark side of decentralisation, Economic Policy 49, 5-70.

Kim, S.H., T.S. Pickton, and S. Gerking (2003), Foreign Direct Investment: Agglomeration Economies and Returns to Promotion Expenditures, The Review of Regional Studies 33: 61-72.

Koppelman, F.S., and C.H. Wen (1998), Alternative Nested Logit Models: Structure Properties and Estimation, Transportation Research 32, 289-298.

Louviere, J.J., D.A. Hensher, and J.Swait (2000), Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Evaluation, Cambridge: Cambridge University Press.

McFadden, D., (1974), Conditional logit analysis of qualitative choice behavior, In: Zarembka P. (Ed.) Frontiers in Econometrics. Academic Press, New York, 105-142.

Schmidheiny, K., and M. Brülhart (2011), On the equivalence of location choice models: Conditional logit, nested logit and Poisson, Journal of Urban Economics 69: 214-222.

UBS (various years), Prices and Earnings - A comparison of purchasing power around the globe, Zurich.

Winkelmann, R. (2008), Econometric Analysis of Count Data, Berlin and Heidelberg: Springer (5^{th} ed.).

A Proof Proposition 1

In a choice consistent nested logit model, adding a constant Δ to all profits (2) and (12), that is $\Pi^*_{sh} = \Pi_{sh} + \Delta$, should not change the probabilities P_{s0} and $P_{sh} = P_{\o s} P_{sh|\o}$ of investing, respectively, in the outside option $h = 0$ or in location $h > 0$. The following demonstrates that the normalisation of proposition 1 fulfills this property whilst this is not the case for the normalisation of SB.

Normalisation of Proposition 1 ($\lambda^o_s = \lambda^{\o}_s = \lambda_s$ and $\varsigma^{\o}_s = 1$ and $\varsigma^o_s = 0$)

Consider first $P_{sh} = P_{\o s} P_{sh|\o}$ of (24).

$$
\begin{aligned}
P_{sh} &= \frac{\exp\left(x'_{sh}\beta\right)\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{(\lambda_s-1)}}{1+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{\lambda_s}} \\
&= \frac{\exp\left(x'_{sh}\beta\right)\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{(\lambda_s-1)}}{\left[\exp(0)\right]^{\lambda_s}+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{\lambda_s}}
\end{aligned}
$$

Adding Δ yields

$$
P^*_{sh} = \frac{\exp\left(x'_{sh}\beta+\Delta\right)\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta+\Delta\right)\right]^{(\lambda_s-1)}}{\left[\exp\left(\Delta\right)\right]^{\lambda_s}+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta+\Delta\right)\right]^{\lambda_s}}.
$$

Now

$$
\begin{aligned}
P^*_{sh} &= \frac{\exp(\Delta)\exp\left(x'_{sh}\beta\right)\left[\exp(\Delta)\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{(\lambda_s-1)}}{\left[\exp(\Delta)\right]^{\lambda_s}+\left[\exp(\Delta)\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{\lambda_s}} \\
&= \frac{\exp(\Delta)\exp\left(x'_{sh}\beta\right)\left[\exp(\Delta)\right]^{(\lambda_s-1)}\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{(\lambda_s-1)}}{\left[\exp(\Delta)\right]^{\lambda_s}+\left[\exp(\Delta)\right]^{\lambda_s}\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{\lambda_s}} \\
&= \frac{\left[\exp(\Delta)\right]^{\lambda_s}\left\{\exp\left(x'_{sh}\beta\right)\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{(\lambda_s-1)}\right\}}{\left[\exp(\Delta)\right]^{\lambda_s}\left\{1+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{\lambda_s}\right\}} \\
&= \frac{\exp\left(x'_{sh}\beta\right)\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{(\lambda_s-1)}}{1+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{\lambda_s}} = P_{sh}
\end{aligned}
$$

and P^*_{sh} equals P_{sh} of (24). Likewise, consider P_{s0} of (25)

$$
P_{s0} = \frac{1}{1+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{\lambda_s}} = \frac{\left[\exp(0)\right]^{\lambda_s}}{\left[\exp(0)\right]^{\lambda_s}+\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x'_{sh}\beta\right)\right]^{\lambda_s}}.
$$

Adding Δ yields

$$
\begin{aligned}
P_{s0}^{*} &= \frac{[\exp(\Delta)]^{\lambda_s}}{[\exp(\Delta)]^{\lambda_s} + \left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta + \Delta\right)\right]^{\lambda_s}} \\
&= \frac{[\exp(\Delta)]^{\lambda_s}}{[\exp(\Delta)]^{\lambda_s} + \left\{\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]\exp(\Delta)\right\}^{\lambda_s}} \\
&= \frac{[\exp(\Delta)]^{\lambda_s}}{[\exp(\Delta)]^{\lambda_s} + [\exp(\Delta)]^{\lambda_s}\left\{\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]\right\}^{\lambda_s}} \\
&= \frac{1}{1 + \left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]^{\lambda_s}} = P_{s0}
\end{aligned}
$$

which is equal to P_{s0} of (25). Hence, the normalisation of Proposition 1 is choice consistent.

Schmidheiny and Brühlhart Normalisation ($\varsigma_s^{\o} = 1$, $\varsigma_s^{o} = 1$, and $\lambda_s^{o} = 1$)

Adding Δ to P_{s0} of (20) yields

$$
P_{s0}^{*} = \frac{\exp(\delta_s + \Delta)}{\left[\exp\left(\delta_s + \Delta\right)\right] + \left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta + \Delta\right)\right]^{\lambda_s^{\o}}}.
$$

and

$$
\begin{aligned}
P_{s0}^{*} &= \frac{\exp(\delta_s)\exp(\Delta)}{\exp(\delta_s)\exp(\Delta) + \left\{\left[\sum_{s=1}^{S}\sum_{h=1}^{H}\exp\left(x_{sh}^{'}\beta\right)\right]\exp(\Delta)\right\}^{\lambda_s^{\o}}} \\
&= \frac{\exp(\delta_s)\exp(\Delta)}{\exp(\delta_s)\exp(\Delta) + \exp(\Delta)^{\lambda_s^{\o}}\left[\left(\sum_{s=1}^{S}\sum_{h=1}^{H}\exp(x_{sh}^{'}\beta)\right)\right]^{\lambda_s^{\o}}} \neq P_{s0}
\end{aligned}
$$

Hence the normalisation of SB is choice inconsistent.

B Derivation of Elasticity

With $E[\widetilde{n}_{sh}^{pcu}] = (E[N^{\o}])^{(\lambda_s - 1)}\exp(x_{sh}^{'}\beta)$ of (26), the elasticity with respect to a (logarithmically transformed) variable $x_{sh,k}$ is

$$
\begin{aligned}
\eta_{sh}^{pcu} &= \frac{\partial E[\widetilde{n}_{sh}^{pcu}]}{\partial x_{sh,k}}\frac{x_{sh,k}}{E[\widetilde{n}_{sh}^{pcu}]} \\
&= \left[(\lambda_s - 1)(E[N^{\o}])^{(\lambda_s - 2)}\exp(.)\frac{\beta_k}{x_{sh,k}}\exp(.) + E[N^{\o}]^{(\lambda_s - 1)}\exp(.)\frac{\beta_k}{x_{sh,k}}\right]\frac{x_{sh,k}}{E[N^{\o}]^{(\lambda_s - 1)}\exp(.)}.
\end{aligned}
$$

Cancelling terms yields

$$
\begin{aligned}
\eta_{sh}^{pcu} &= (\lambda_s - 1)\underbrace{(E[N^{\o}])^{-1}\exp(x_{sh}^{'}\beta)}_{=P_{sh|\o}\text{ according to (14)}}\beta_k + \beta_k \\
&= [1 + (\lambda_s - 1)P_{sh|\o}]\beta_k = [1 - (1 - \lambda_s)P_{sh|\o}]\beta_k.
\end{aligned}
$$

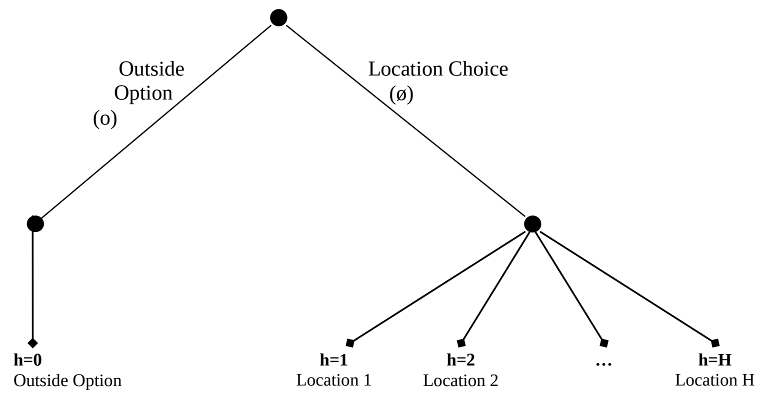
C Data Appendix

Description of the Data Set

Variable	***Unit***	***Description***	***Source***
Dependent Variables:			
$l_{i,sh}$	Nominal	For each cross-border acquisition i from source s, this indicates whether host h has been chosen, in which case $l_{i,sh} = 1$, or another location has been chosen as host, in which case $l_{i,sh} = 0$.	Compiled from SDC Platinum of Thomson Financial.
n_{sh}	Count	Number of cross-border acquisition deals between source s and host h during 2005 to 2009 period.	Compiled from SDC Platinum of Thomson Financial.
Independent Variables:			
Common border (BORDER)	Nominal	Dummy variable for countries sharing a common land border.	Own calculations.
Corruption (CORRUPT)	Index	Corruption index on a scale from 10 to 90. Original values have been reversed such that higher values mean more corruption.	Heritage Foundation.
Distance (DIST)	1,000 Km.	Great circular distance between capital city of the source and host.	Own calculations.
Gross Domestic Product (GDP)	Bn. US$	Real gross domestic product in US$ with base year 2000 of the host h.	World Development Indicators.
Investment Freedom (INVFR)	Index	Index of freedom of investment referring to whether there is an FDI code that defines the country's investment laws and procedures; whether the government encourages FDI through fair and equitable treatment of investors; whether there are restrictions on access to foreign exchange; whether foreign firms are treated the same as domestic firms under the law; whether the government imposes restrictions on payments, transfers, and capital transactions; and whether specific industries are closed to FDI.	Heritage Foundation.
Common Language (LANG)	Nominal	Dummy variable for countries sharing a common official language.	Compiled from CIA World Factbook.
Corporate Tax (TAX)	Per cent	Effective average tax rate (EATR) on corporate income on outbound FDI between the source and host (overall case). The EATR is the proportion of profit from an investment taken in tax accounting for capital depreciation, tax allowances, withholding taxes, and international tax relief.	Devereux *et al.* (2009).
Trade Freedom (TRADEFR)	Index	Index of freedom of international trade (tariff and non-tariff barriers) on a scale from 10 to 90.	Heritage Foundation.
Wage (WAGE)	Index (Zurich =100)	Wage level in the host. Wages are measured by an index referring to the hourly gross income of 13 comparable professions (product managers, department heads, engineers, primary school teachers, bus drivers, car mechanics, building labourers, industrial workers, cooks, bank credit officers, personal assistants, sales assistants, factory workers) as paid in the capital city or the financial centre of a country.	UBS, Prices and Earnings.

D Figures

Figure 1: Location Choice and Outside Options



E Tables

Table 1: Number and Probability of CBAs between 25 EU Countries (2005 to 2009)

	Number of CBAs between 25 EU Countries (2005 to 2009) with Host																								
Source	AUT	BEL	BGR	CZE	DEU	DNK	ESP	EST	FIN	FRA	GBR	GRC	HUN	IRL	ITA	LTU	LUX	LVA	NLD	POL	PRT	ROM	SVK	SVN	SWE
AUT		7	22	26	162	3	5	2	1	13	13	2	23	3	20	0	1	2	12	24	0	29	12	9	10
BEL	7		3	6	39	5	21	3	7	99	28	0	11	2	16	1	12	1	69	9	7	5	3	0	5
BGR	0	0		0	1	0	0	0	0	0	2	1	0	0	0	1	0	1	0	0	0	3	0	0	0
CZE	2	0	5		12	0	0	0	0	0	3	0	4	0	0	0	1	0	0	16	0	11	12	2	0
DEU	143	46	20	55		51	63	1	28	146	182	11	33	19	89	6	11	2	91	42	10	11	9	1	55
DNK	3	7	0	3	56		16	4	23	14	26	1	0	3	8	5	1	2	20	18	2	7	2	0	122
ESP	5	8	2	7	41	5		3	8	85	53	4	3	2	50	1	1	0	22	17	63	5	1	0	5
EST	0	0	1	0	0	1	0		9	1	0	3	2	0	0	11	0	11	1	2	0	0	0	0	0
FIN	5	3	1	6	32	25	8	19		11	22	0	3	0	13	13	0	7	10	21	0	2	6	1	114
FRA	13	133	10	31	159	24	139	1	7		166	15	22	19	110	5	18	6	71	35	22	17	7	2	29
GBR	15	48	13	24	287	45	111	3	40	202		7	21	102	89	4	11	0	105	41	21	23	7	2	97
GRC	2	3	19	0	5	1	6	0	0	1	2		1	0	5	0	0	0	1	2	0	20	0	0	1
HUN	1	0	2	1	2	0	1	0	1	1	1	0		0	2	0	0	0	3	5	2	12	0	0	0
IRL	3	8	4	5	21	6	6	1	0	10	176	0	1		2	0	1	2	17	1	2	1	1	0	6
ITA	13	10	6	8	68	4	57	0	6	67	58	6	8	3		0	9	3	21	11	5	7	1	4	6
LTU	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0		0	8	2	3	0	0	1	0	0
LUX	2	17	8	6	54	3	11	4	2	47	24	3	1	2	40	1		0	19	6	3	3	3	0	6
LVA	0	1	0	0	2	0	0	2	0	0	1	0	0	0	0	6	0		0	0	0	0	0	0	2
NLD	14	129	9	28	181	22	61	2	17	73	141	2	15	9	59	3	7	3		30	7	17	5	3	34
POL	4	1	1	19	8	3	8	0	0	1	1	0	6	1	0	9	1	0	6		0	13	3	0	2
PRT	1	1	1	2	10	1	45	0	0	10	8	2	0	2	2	0	0	0	2	4		6	1	0	2
ROM	1	0	2	1	0	0	1	0	0	1	1	0	2	0	3	0	0	0	0	0	0		0	0	0
SVK	0	0	0	12	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0		1	0
SVN	1	0	2	1	2	0	1	0	0	1	1	1	0	0	1	0	0	0	1	0	0	0	0		0
SWE	11	18	8	11	78	121	31	20	140	38	105	2	1	4	17	17	3	10	41	19	3	3	3	3	

	Probability $P_{sh\|\o}$ of CBA with Host																								
Source	AUT	BEL	BGR	CZE	DEU	DNK	ESP	EST	FIN	FRA	GBR	GRC	HUN	IRL	ITA	LTU	LUX	LVA	NLD	POL	PRT	ROM	SVK	SVN	SWE
AUT		3.3%	1.6%	3.9%	34.4%	1.3%	2.7%	0.4%	1.1%	5.5%	6.6%	0.8%	5.8%	0.9%	7.7%	0.7%	0.8%	0.5%	3.4%	3.0%	1.1%	1.2%	9.1%	1.9%	2.3%
BEL	1.4%		0.9%	1.1%	11.8%	1.1%	2.7%	0.3%	0.8%	31.5%	11.8%	0.5%	1.2%	1.1%	3.0%	0.4%	2.9%	0.3%	20.6%	1.6%	1.1%	0.7%	0.9%	0.5%	1.8%
BGR	3.9%	2.5%		2.1%	12.8%	1.9%	5.2%	0.7%	1.8%	8.8%	10.4%	5.1%	4.0%	1.4%	9.5%	1.2%	0.6%	0.8%	5.1%	4.7%	2.1%	7.8%	2.6%	1.3%	3.8%
CZE	6.7%	2.2%	1.5%		30.9%	1.8%	3.1%	0.5%	1.3%	6.8%	8.4%	0.8%	4.4%	1.0%	4.7%	0.8%	0.5%	0.6%	4.6%	8.7%	1.3%	1.2%	4.2%	1.0%	2.9%
DEU	11.1%	4.6%	1.6%	5.7%		4.9%	3.7%	0.6%	1.7%	13.7%	10.8%	0.9%	2.6%	1.3%	5.2%	1.1%	2.2%	0.8%	10.3%	7.0%	1.6%	1.4%	2.2%	0.9%	4.0%
DNK	2.7%	2.7%	1.6%	2.2%	32.0%		3.7%	0.8%	2.0%	8.0%	11.4%	0.9%	2.3%	1.5%	4.7%	1.1%	0.6%	0.9%	6.5%	3.5%	1.6%	1.3%	1.8%	0.8%	5.4%
ESP	2.6%	3.0%	1.9%	1.7%	10.9%	1.7%		0.6%	1.5%	20.5%	14.4%	1.2%	2.3%	2.1%	7.6%	0.9%	0.7%	0.7%	5.8%	2.9%	9.8%	1.5%	1.7%	0.9%	3.3%
EST	2.8%	2.4%	1.6%	1.9%	14.8%	2.6%	4.5%		9.7%	8.4%	11.2%	1.1%	2.4%	1.4%	5.6%	1.8%	0.6%	3.0%	5.5%	3.4%	1.8%	1.6%	1.8%	0.8%	9.3%
FIN	2.5%	2.2%	1.7%	1.7%	12.2%	2.3%	3.8%	3.2%		7.1%	9.7%	1.0%	2.3%	1.4%	4.8%	1.6%	0.5%	1.6%	4.8%	3.2%	1.7%	1.6%	1.7%	0.7%	26.8%
FRA	2.1%	14.8%	1.3%	1.5%	16.6%	1.5%	8.5%	0.4%	1.2%		17.9%	0.8%	1.8%	1.7%	8.7%	0.7%	2.9%	0.5%	6.8%	2.3%	2.0%	1.0%	1.4%	0.7%	2.7%
GBR	2.6%	5.9%	1.3%	1.9%	14.0%	2.2%	6.4%	0.6%	1.7%	19.1%		1.0%	2.0%	8.5%	6.3%	0.8%	1.1%	0.6%	11.4%	2.8%	2.6%	1.2%	1.5%	0.8%	3.9%
GRC	3.5%	2.7%	6.9%	2.0%	13.0%	1.9%	6.0%	0.7%	2.0%	9.5%	11.6%		3.3%	1.3%	10.6%	1.1%	0.6%	0.7%	5.4%	4.1%	2.5%	3.1%	2.2%	1.2%	4.0%
HUN	9.7%	2.3%	2.9%	2.6%	13.7%	1.8%	4.0%	0.6%	1.6%	7.7%	9.4%	1.4%		1.2%	7.2%	1.2%	0.6%	0.8%	4.7%	7.6%	1.7%	3.9%	7.5%	2.5%	3.4%
IRL	1.6%	2.4%	1.0%	1.2%	8.0%	1.4%	4.2%	0.4%	1.2%	8.5%	48.9%	0.7%	1.3%		3.8%	0.6%	0.5%	0.4%	5.2%	1.9%	1.8%	0.9%	1.0%	0.5%	2.6%
ITA	6.1%	2.9%	3.0%	2.2%	12.8%	1.8%	6.5%	0.6%	1.7%	17.7%	12.1%	1.9%	3.4%	1.6%		1.0%	0.7%	0.7%	5.6%	4.7%	2.6%	2.1%	2.4%	2.5%	3.5%
LTU	3.4%	2.5%	2.4%	2.4%	15.7%	2.5%	4.3%	1.3%	3.3%	8.1%	10.8%	1.3%	3.3%	1.5%	6.0%		0.6%	4.0%	5.4%	7.4%	1.9%	2.1%	2.3%	1.0%	6.3%
LUX	2.9%	12.5%	0.8%	1.1%	24.3%	1.1%	2.5%	0.3%	0.8%	26.4%	9.1%	0.5%	1.2%	0.9%	3.0%	0.4%		0.3%	5.4%	1.6%	1.0%	0.7%	0.9%	0.5%	1.7%
LVA	3.1%	2.5%	2.2%	2.2%	15.5%	2.7%	4.3%	3.0%	4.2%	8.1%	10.9%	1.2%	2.9%	1.6%	5.7%	5.2%	0.6%		5.5%	4.1%	1.9%	1.9%	2.1%	0.9%	7.8%
NLD	2.2%	15.9%	1.3%	1.7%	20.6%	2.0%	4.0%	0.5%	1.3%	11.2%	18.0%	0.8%	1.8%	1.7%	4.5%	0.7%	1.0%	0.6%		2.5%	1.6%	1.1%	1.4%	0.7%	3.1%
POL	3.6%	2.2%	2.0%	4.2%	26.9%	2.3%	3.7%	0.9%	2.2%	7.3%	9.5%	1.1%	3.3%	1.2%	5.5%	1.9%	0.5%	2.0%	4.9%		1.6%	3.2%	4.2%	1.0%	4.7%
PRT	2.4%	2.8%	1.7%	1.6%	10.4%	1.7%	22.5%	0.6%	1.6%	11.0%	13.7%	1.2%	2.2%	2.1%	7.0%	0.9%	0.6%	0.7%	5.5%	2.8%		1.4%	1.6%	0.8%	3.3%
ROM	3.4%	2.3%	8.8%	2.0%	12.4%	1.8%	4.6%	0.7%	1.9%	7.8%	10.0%	2.3%	6.7%	1.4%	7.7%	1.3%	0.5%	0.9%	4.8%	7.6%	1.9%		4.2%	1.1%	3.7%
SVK	9.6%	1.9%	1.8%	4.3%	12.2%	1.5%	3.2%	0.5%	1.3%	6.3%	7.7%	1.0%	7.9%	1.0%	5.4%	0.9%	0.5%	0.6%	3.9%	20.6%	1.3%	2.5%		1.3%	2.7%
SVN	8.4%	2.4%	2.2%	2.5%	12.8%	1.6%	4.2%	0.5%	1.4%	8.4%	9.7%	1.2%	6.4%	1.2%	13.9%	0.9%	0.6%	0.6%	4.8%	6.5%	1.7%	1.7%	3.1%		3.0%
SWE	2.9%	2.6%	1.9%	2.1%	15.6%	3.2%	4.4%	1.6%	14.4%	8.5%	11.9%	1.1%	2.6%	1.7%	5.5%	1.7%	0.6%	1.6%	6.1%	3.7%	1.9%	1.6%	1.9%	0.9%	

Table 2: Group Effects α_s and Similarity Parameter

Source	α_s (1)	λ_s (2)	Source	α_s (3)	λ_s (4)
AUT	0.050	0.75	IRL	0.047	0.74
BEL	0.035	0.72	ITA	0.098	0.80
BGR	0.002	0.49	LTU	0.005	0.55
CZE	0.010	0.61	LUX	0.025	0.69
DEU	0.199	0.86	LVA	0.003	0.52
DNK	0.063	0.77	NLD	0.144	0.84
ESP	0.108	0.81	POL	0.017	0.66
EST	0.013	0.63	PRT	0.030	0.70
FIN	0.074	0.78	ROM	0.003	0.51
FRA	0.179	0.85	SVK	0.003	0.49
GBR	0.311	0.90	SVN	0.002	0.48
GRC	0.022	0.68	SWE	0.174	0.85
HUN	0.007	0.57			