On Nesting Location Choice Models Correctly:
A Reply to Herger and McCorriston (Economics Letters, 2013)*

Kurt Schmidheiny† Marius Brühlhart§
University of Basel University of Lausanne

March 3, 2015

Abstract
As shown by Guimaraes, Figueiredo and Woodward (2003), a particular class of conditional
logit models yield identical parameter estimates to a Poisson count data model. In Schmid-
heiny and Brühlhart (2011), we have pointed out that the conditional logit model and the
Poisson model can be seen as polar cases of a continuum of intermediate cases which emerge
from a random utility nested logit model. Herger and McCorriston (2013) have proposed
an alternative nested logit model. In this note, we show that their paper misrepresents our
model, that their alternative specification does not in fact nest the Poisson model, and that
their analysis contains a number of formal errors.

Highlights:
• The nested logit model in Schmidheiny and Brühlhart (2011) is a random utility model
  that nests the conditional logit and Poisson models.
• The nested logit model in Herger and McCorriston (2013) is a restricted version of
  Schmidheiny and Brühlhart (2011) that does not nest the Poisson model.
• The proposed estimation strategy in Herger and McCorriston (2013) is incorrect.

JEL Classification: C25, R3
Keywords: location choice, conditional logit, nested logit, Poisson count model

*We thank Jörg Kalbfuss for carefully reading the manuscript. We gratefully acknowledge financial support
from the Swiss National Science Foundation (Sinergia Grants 130648 and 147668; and NCCR “Trade Regulation”).
†Wirtschaftswissenschaftliche Fakultät, Universität Basel, Peter Merian-Weg 6 4125 Basel, Switzerland; e-mail:
kurt.schmidheiny@unibas.ch. Also affiliated with the the Centre for Economic Policy Research (CEPR) and
CESifo.
§Department of Economics, Faculty of Business and Economics, University of Lausanne, CH - 1015 Lausanne,
Switzerland. E-mail: Marius.Bruhlhart@unil.ch. Web page: www.hec.unil.ch/mbruhar. Also affiliated with
the Centre for Economic Policy Research (CEPR).
1 Introduction

Location choice models explain the spatial distribution of agents such as firms or households as the outcome of individual location decisions. The most popular representation of such choices is McFadden’s (1974) conditional logit model. Guimaraes, Figueiredo and Woodward (2003) have shown that a particular class of conditional choice models yield identical parameter estimates to a Poisson count data model. Schmidheiny and Brülhart (2011, henceforth SB) show that, while the two models are indeed observationally equivalent in cross-section data, they are distinctively different descriptions of the world and imply different marginal effects and elasticities. SB also show that the elasticities of the conditional logit model and the Poisson model can be seen as polar cases of a continuum of intermediate cases which emerge from a random utility nested logit model (RUNL). Herger and McCorriston (2013, henceforth HM) propose an alternative nested logit model. They claim that their model, unlike the SB RUNL, allows the “dissimilarity parameter” to be identified in cross-section data, thus making it possible to estimate how closely the data correspond to the conditional logit or Poisson models. In this note, we show that HM wrongly represent SB’s RUNL and that their proposed alternative specification is a restricted version of SB which does not nest the Poisson model.

2 Conditional Logit and Poisson

Using the notation of HM, the basic framework in Guimaraes, Figueiredo and Woodward (2003) is the following

- Firms $i = 1, ..., N$ choose among $h = 1, ..., H$ locations. $h$ are e.g. states within a country.
- There are $s = 1, ..., S$ different types of firms. $s$ are e.g. industrial sectors or country of origin for foreign firms.
- The number of firms $n_{sh}$ of each type $s$ is observed in each location $h$. $n_{sh}$ is used to denote both the observed number and the random variable in the data generating process.
- A vector of observable explanatory variables is denoted $x_{sh}$, characterizing locations $h$ and potentially also differing across types $s$.

The location choice of an individual firm $i$ is assumed to correspond to a conditional logit model. Thus, the profit function

$$\pi_{ish} = x_{sh}'\beta + \varepsilon_{ish}$$

(1)

consists of a type-specific deterministic part $x_{sh}'\beta$ and a stochastic part $\varepsilon_{ish}$, which has zero mean and is known to the firm but unknown to the econometrician. Firms choose the location which yields the highest profit.

Comment 1 The expected profit function $E(\pi_{ish})$ in the conditional logit model does not contain an error term as stated in HM eq. (1) and (2). It is not useful to include a constant ($\delta_s$ in HM) or type-specific fixed effect in (1) as stated in HM eq. (2), because this type-specific constant is not identified in the conditional logit model.

The conditional logit model contains the assumption that the stochastic term $\varepsilon_{ish}$ is independent across $i$ and $h$ and follows an extreme value type 1 distribution. The conditional logit model implies the following choice probabilities:

$$P_{sh}^{CL} = \frac{e^{x_{sh}'\beta}}{\sum_{g=1}^{H} e^{x_{sg}'\beta}}.$$ 

(2)
Comment 2 The choice probabilities in (2) are conditional on the type s. The denominator in eq. 2 does not sum over the firm types s as stated in HM eq. (3). The double summation in HM eq. (3) implies that the firm’s type s is a choice of the firm rather than a characteristic, which is not what the basic framework assumes and which contradicts HM’s own decision rule in HM eq. (1).

The conditional logit model only predicts probabilities (relative numbers) and not absolute numbers of location choices, as it does not explicitly model the total number \( n_s \) of firms of each type. The total number \( n_s \) of firms of type s is often implicitly assumed as given in the conditional logit model. With exogenous \( n_s \), the expected absolute number of choices is

\[
E(n_{sh}^{CL}) = n_s P_{sh}. \tag{3}
\]

SB derive the effect of a change in the \( k \)-th characteristic \( x_{shk} \) on the expected number of firms of type s in location h as the following semi-elasticity:

\[
\eta_{sh}^{CL} = \frac{\partial \log E(n_{sh}^{CL})}{\partial x_{shk}} = \frac{\partial E(n_{sh}^{CL})}{\partial x_{shk}} \frac{1}{E(n_{sh}^{CL})} = (1 - P_{sh}) \beta_k, \tag{4}
\]

assuming an exogenous number \( n_s \) of firms of each type.

Comment 3 \((1 - P_{sh}) \beta_k\) equals \(\partial E(n_{sh}^{CL})/\partial x_{shk}\)·\([1/E(n_{sh}^{CL})]\) and not \(\partial E(n_{sh}^{CL})/\partial x_{shk}\)·\([x_{shk}/E(n_{sh}^{CL})]\) as stated in HM eq. (10). This expression is a semi-elasticity if \(x_{shk}\) is defined in levels and an elasticity if \(x_{shk}\) is defined in logs.\(^1\)

Guimaraes, Figueiredo and Woodward (2003) introduce a Poisson count data model which yields identical parameter estimates for the slope coefficients \(\beta\). The Poisson approach directly models the expected absolute number of choices:

\[
E(n_{sh}^{PC}) = e^{\alpha_s + x_{sh} \beta}. \tag{5}
\]

The Poisson model is a distinctively different data generating process from the conditional logit model and hence a different description of the world.

Comment 4 The type-specific constants \(\alpha_s\) in the Poisson model are unrelated to fixed effects in the profit function (1) of the conditional logit model, contrary to the implied assertion by HM eq. (2) and (5).\(^2\)

In the Poisson model, the effect of a change in the \(k\)-th local characteristic \(x_{shk}\) on the expected number of firms of type s in location h as semi-elasticity is

\[
\eta_{sh}^{PC} = \frac{\partial \log E(n_{sh}^{PC})}{\partial x_{shk}} = \frac{\partial E(n_{sh}^{PC})}{\partial x_{shk}} \frac{1}{E(n_{sh}^{PC})} = \beta_k. \tag{6}
\]

Comment 5 \(\beta_k\) equals \(\partial E(n_{sh}^{PC})/\partial x_{shk}\)·\([1/E(n_{sh}^{PC})]\) and not \(\partial E(n_{sh}^{PC})/\partial x_{shk}\)·\([x_{shk}/E(n_{sh}^{PC})]\) as stated in HM eq. (9). This expression is a semi-elasticity if \(x_{shk}\) is defined in levels and an elasticity if \(x_{shk}\) is defined in logs.\(^3\)

\(^1\)HM define \(x_{shk}\) as reported in logs. We see no value added in restricting \(x_{shk}\) to strictly positive values.

\(^2\)We find the derivation of the Poisson model in HM to be formally incorrect and its interpretation to be misleading. While the definition of the Poisson model in HM eq. (5) is accurate, it is not the result of multiplying the CL probability in HM eq. (3) by its denominator. The constant term \(\alpha_s = \exp(\delta_s)\) in HM eq. (5) can therefore not be related to the non-identified constant \(\delta_s\) in HM eq. (2). We also find the interpretation in the first line after HM eq. (7) misleading, because \(E(n_{sh}^{PC}) = \sum_{h=1}^{H} e^{\alpha_s + x_{sh} \beta}\) is not the expected number of location choices in the Poisson model (which instead is \(E(n_{sh}^{PC}) = E(\sum_{h=1}^{H} n_{sh}^{PC}) = \sum_{h=1}^{H} e^{\alpha_s + x_{sh} \beta}\)). Moreover, while the ML estimator of \(\alpha_s\) in HM eq. (7) is correct, it cannot be interpreted as the “discrepancy” between expected and observed numbers of observations.
3 The RUNL

In SB, we have proposed the following random utility nested logit (RUNL) model. The profit of a firm $i$ of type $s$ in location $h = 1, ..., H$ is given by

$$\pi_{ish} = x_{ish}' \gamma_s + \epsilon_{ish} = x_{ish}' \beta \lambda_s + \epsilon_{ish},$$ (7)

where the reparametrization $\gamma_s = \beta \lambda_s$ is described in SB, p. 216. This reparametrization, in combination with the assumptions that $\beta$ is constant across types $s$ and $\gamma_s$ is type-specific, formally allows us to relate the coefficient vector $\beta$ to both the CL and the Poisson model. The RUNL model of SB moreover considers an outside option $h = 0$, which can be interpreted as locating in another country or not starting up a company. The outside option yields profit

$$\pi_{iso} = \delta_s + \epsilon_{iso},$$ (8)

It is assumed in the RUNL model that the stochastic term $\epsilon_{ish}$ follows a special type of generalized extreme value (GEV) distribution (see McFadden, 1978) such that all locations $h = 1, ..., H$ belong to one nest and the outside option $h = 0$ belongs to another (degenerate) nest. Following Amemiya (1985, p. 303, eq. 9.3.60) the choice probabilities are then expressed as:

$$P^{RUNL}_{ish} = (1 - P^{RUNL}_{iso}) \cdot P^{RUNL}_{ish|\emptyset}$$ for all $h > 0,$ (9)

$$P^{RUNL}_{iso} = \frac{e^{\delta_s} \left( \sum_{g=1}^{H} e^{x_{sg}' \gamma_s / \lambda_s} \right)^{\lambda_s}}{e^{\delta_s} \left( \sum_{g=1}^{H} e^{x_{sg}' \beta / \lambda_s} \right)^{\lambda_s}},$$ (10)

$$P^{RUNL}_{ish|\emptyset} = \frac{e^{x_{ish}' \gamma_s / \lambda_s}}{\sum_{g=1}^{H} e^{x_{sg}' \gamma_s / \lambda_s}} = \frac{e^{x_{ish}' \beta / \lambda_s}}{\sum_{g=1}^{H} e^{x_{sg}' \beta / \lambda_s}},$$ (11)

where $P^{RUNL}_{ish|\emptyset}$ is the probability of choosing location $h > 0$ conditional on not choosing the alternative. The parameter $\lambda_s$ can be interpreted as $\lambda_s = \sqrt{1 - \rho_s}$, where $\rho_s$ is the correlation of the stochastic term $\epsilon_{ish}$ within the choices $h > 0$. Note that this version of the RUNL is fully general; other versions of the nested logit model which are consistent with a random utility model are just reparametrizations of this specification.

Comment 6 Adding a constant $\Delta$ to the profit functions of all choice options $h = 0, 1, ..., H$ does not alter the choice probabilities in the SB RUNL. HM p. 290 wrongly state the opposite, implying that the SB nested logit is not a RUNL.

The nested logit (NL) model in Hunt (2000) encompasses specifications which are consistent with a random utility model and others which are not:

$$P^{NL}_{ish} = (1 - P^{NL}_{iso}) \cdot P^{NL}_{ish|\emptyset}$$ for all $h > 0,$ (12)

$$P^{NL}_{iso} = \frac{\left( e^{\delta_s \varsigma^o} \right)^{\lambda_s^o}}{\left( e^{\delta_s \varsigma^o} \right)^{\lambda_s^o} + \left( \sum_{g=1}^{H} e^{x_{sg}' \beta \varsigma^o / \varsigma^o} \right)^{\lambda_s^o}},$$ (13)
The RUNL in SB is a special case of the Hunt (2000) NL model with $\beta_s = \beta$, $\lambda_o^s = \lambda_s$, $\lambda_i^o = 1$, $\zeta_s^o = 1$ and $\zeta_i^o = 1$.

In the SB RUNL, the expected number of firms of type $s$ in location $h$ is

$$E(n_{sh}^{\text{RUNL}}) = (n_{so+sh})^\text{RUNL}.$$  \hfill (15)

where $n_{so+sh}$ is the exogenous total number of firms of type $s$ which choose either $h > 0$ or $h = 0$. The total, $n_{so+sh}$, is typically unobservable. The semi-elasticity of a change in the $k$-th local characteristic $x_{shk}$ on the expected number of firms of type $s$ in location $h > 0$ is

$$\eta_{sh}^\text{RUNL} = \frac{\partial \log E(n_{sh}^{\text{RUNL}})}{\partial x_{shk}} = [1 - P_{sh|\phi}^{\text{RUNL}}(1 - \lambda_s P_{so}^{\text{RUNL}})] \beta_k.$$  \hfill (16)

The RUNL elasticity $\eta_{sh}^\text{RUNL}$ equals the CL elasticity if $\lambda_s = 0$ and the Poisson elasticity if $\lambda_s = 1$ and $\delta_s \to \infty$, hence $P_{so}^{\text{RUNL}} = 1$. The SB RUNL therefore nests the conditional logit model, the Poisson model and the continuum of cases in-between.

### 4 The Herger and McCorriston RUNL

The NL model proposed in HM is a special case of the Hunt (2000) NL model with $\beta_s = \beta$, $\lambda_o^s = \lambda_s$, $\lambda_i^o = \lambda_s$, $\zeta_s^o = 1$, $\zeta_i^o = 0$ and/or $\delta_s^o = 0$, resulting in (see HM eq. 20 and 21)

$$P_{sh}^{\text{HM}} = (1 - P_{so}) \cdot P_{sh|\phi} = \frac{e^{x_{sh}^\beta} \left( \sum_{g=1}^{H} e^{x_{sg}^\beta} \right)^{\lambda_s - 1}}{1 + \left( \sum_{g=1}^{H} e^{x_{sg}^\beta} \right)^{\lambda_s}} \quad \text{for all } h > 0,$$  \hfill (17)

$$P_{so}^{\text{HM}} = \frac{1}{1 + \left( \sum_{g=1}^{H} e^{x_{sg}^\beta} \right)^{\lambda_s}} = \frac{1}{1 + \left( \sum_{g=1}^{H} e^{x_{sg}^\beta} \right)^{\lambda_s}}, \quad \text{and}$$  \hfill (18)

$$P_{sh|\phi}^{\text{HM}} = \frac{e^{x_{sh}^\beta}}{\sum_{g=1}^{H} e^{x_{sg}^\beta}}.$$  \hfill (19)

Comparing (10) and (18) shows that the HM specification is a restricted version of the SB RUNL with the implicit restriction $\delta_s = 0$ for all $s$.

**Comment 7** The HM RUNL is a restricted version of the fully general SB RUNL. There is no parameter to control for the value of the outside option in general, and the value of the outside option is assumed to be identical across types $s$.

The semi-elasticity in the HM RUNL is identical to the one in the SB RUNL in eq. (16) but with $P_{sh|\phi}^{\text{HM}}$ and $P_{sh|\phi}^{\text{RUNL}}$ in place of $P_{sh|\phi}^{\text{RUNL}}$ and $P_{sh|\phi}^{\text{RUNL}}^{\text{RUNL}}$, respectively:

$$\eta_{sh}^{\text{HM}} = \frac{\partial \log E(n_{sh}^{\text{HM}})}{\partial x_{shk}} = [1 - P_{sh|\phi}^{\text{HM}}(1 - \lambda_s P_{so}^{\text{HM}})] \beta_k.$$  \hfill (20)

This semi-elasticity cannot be equal to the one of the Poisson model (eq. 6), because $P_{so}^{\text{HM}} < 1$ (strictly) in eq. (18) and hence $1 - \lambda_s P_{so}^{\text{HM}} \neq 0$.

---

\footnote{Setting $\zeta_i^o = 0$ has the same effect as setting $\delta_s^o = 0$ in eq. (13).}
Comment 8 The HM RUNL model does not nest the Poisson model. It cannot in fact be used to study the main question at hand: the equivalence of the CL and Poisson models. The elasticity in HM eq. (24) and Appendix B is incorrect.\textsuperscript{7}

5 Estimation

Guimaraes, Figueiredo and Woodward (2003) show that the conditional logit model (2) and the Poisson model (5) have identical concentrated likelihood functions (up to an additive constant) for the parameter vector \( \beta \) when the estimation is based on one cross-section of observations of \( n_{sh}, h = 1, ..., H \). SB show in Appendix A.3 that this is also true for the coefficient vector \( \beta \) in the RUNL model, assuming \( \lambda_s = \lambda \) (see also footnote 4). It is straightforward to show that this result also holds for type-specific \( \lambda_s \).\textsuperscript{8} Based on data for \( h > 0 \), the three models are therefore observationally equivalent while the implied marginal effects and elasticities differ. Note that the nested logit parameters \( \delta_s \) and \( \lambda_s \) in the SB RUNL are not identified without observing \( n_{so} \). If \( n_{so} \) were observed, either \( \lambda_s \) or \( \delta_s \) would be identified but not independently. In practice, \( n_{so} \) is typically unknown. For example, the number of potential firms that could start up but do not materialize cannot be observed. So, in practice, neither \( \lambda_s \) nor \( \delta_s \) are identified.

Unlike in the SB RUNL, \( \lambda_s \) is identified in the HM RUNL for the (unlikely) case that \( n_{so} \) were observed. So, as usual, parameter restrictions (here \( \delta_s = 0 \)) help identify structural parameters. But, as shown in Comment 8, the restriction \( \delta_s = 0 \) is not a useful one in this framework. The ML estimator of \( \lambda_s \) with the HM restriction \( \delta_s = 0 \) is

\[
\hat{\lambda}_s = \frac{\log (n_{so}/n_{so})}{\log \left( \sum_{h=1}^{H} \exp(x'_{sh}\hat{\beta}) \right)}, \tag{21}
\]

where \( n_{so} = \sum_{h=1}^{H} n_{sh} \) and \( \hat{\beta} \) is the ML estimator of \( \beta \) (see the derivation in the Appendix). HM propose a different estimator for \( \lambda_s \) in HM eq. (23). Substituting HM eq. (7) and eq. (21) into HM eq. (23) yields:\textsuperscript{9}

\[
\hat{\lambda}_s = \frac{\log (n_{so})}{\log \left( \sum_{h=1}^{H} \exp(x'_{sh}\hat{\beta}) \right)}, \tag{22}
\]

Comment 9 The estimator of \( \lambda_s \) in HM eq. (23) implicitly assumes that the number of firms choosing the outside option equals one, \( n_{so} = 1 \). We cannot see a justification for this assumption in a typical application.

Br imgUrlart and Schmidheiny (2015) show how a panel of observations of \( n_{sh} \) over multiple time periods \( t \) allows the researcher to identify a single parameter \( \rho_s \) which is a function of both \( \delta_s \) and \( \lambda_s \) without the need to observe \( n_{so} \).

\textsuperscript{7}The correct elasticity in the HM model is \( \partial \log E(n'_{sh})/\partial \log x_{sh} = [1 - P_{sh,h}(1 - x_{sh})]\delta_h \cdot x_{sh} \), which contains the term \( P_{sh,h} \). The definitions \( E(N^a) \equiv \sum_{h=1}^{H} e^{x'_{sh}\beta} \) (HM, eq. 20), and \( E(N) \equiv 1 + E(N^a) \) (HM, eq. 22) are not meaningful, and we see no formal basis for HM eq. (23).

\textsuperscript{8}Sketch of proof: Solve the first order condition \( \partial \log L(\beta, \delta, \lambda) / \partial \delta_s \) for \( \delta_s \). Plug the resulting \( \delta_s \) into \( \log L(\beta, \delta, \lambda) \) for all \( s \). \( \lambda_s \) cancels out in the resulting concentrated likelihood function \( \log L(\beta) \).

\textsuperscript{9}HM equation (21) implicitly defines \( E(n_{so}) = \sum_{s=1}^{S} \sum_{h=1}^{H} \exp(x'_{sh}\beta) \). According to our comment 2 this should actually read \( E(n_{so}) = \sum_{h=1}^{H} \exp(x'_{sh}\beta) \). HM eq. (7) estimates \( \hat{\lambda}_s = n_{so}/ \sum_{h=1}^{H} \exp(x'_{sh}\beta) \). Plugging these two elements in HM eq. (23) yields our eq. (22).
6 Appendix

The log likelihood function of the HM RUNL is

$$\log L(\beta, \lambda) = \sum_{s=0}^{S} \left[ n_{so} \log(P_{so}^{HM}) + \sum_{h=1}^{H} n_{sh} \log(P_{sh}^{HM}) \right]$$

$$= \sum_{s=0}^{S} \left[ n_{so} \log \left( \frac{1}{1 + e^{IV_s \lambda_s}} \right) + \sum_{h=1}^{H} n_{sh} \log \left( \frac{e^{x'_{sh} \beta} e^{IV_s (\lambda_s - 1)}}{1 + e^{IV_s \lambda_s}} \right) \right]$$

where $IV_s = \log \sum_{g=1}^{H} e^{x'_{sg} \beta}$ is often called the inclusive value.

The first order condition with respect to $\lambda_s$ is for all $s$

$$\frac{\partial \log L(\beta, \lambda)}{\partial \lambda_s} = -n_{so} IV_s e^{IV_s \lambda_s} + \sum_{h=1}^{H} n_{sh} \left( IV_s - \frac{IV_s e^{IV_s \lambda_s}}{1 + e^{IV_s \lambda_s}} \right) = 0.$$ 

Solving the first order condition for $\lambda_s$ yields

$$\lambda_s = \frac{\log (n_{so}/n_{so})}{IV_s}.$$ 

Plugging $\lambda_s$ into the log likelihood function gives the concentrated log likelihood function

$$\log L(\beta) = \sum_{s=1}^{S} \left[ \sum_{h=1}^{H} n_{sh} x'_{sh} \beta - n_{so} \log \left( \sum_{h=1}^{H} e^{x'_{sh} \beta} \right) \right]$$

$$+ n_{so} \log(n_{so}) + n_{so} \log(n_{so}) - (n_{so} + n_{so}) \log(n_{so} + n_{so})$$

which is a function of $\beta$ only and which is up to a constant identical to the concentrated likelihood functions of the conditional logit, the Poisson and the SB RUNL models. Hence all four models yield identical ML estimates for $\beta$.

References


