Multinomial Logit Models for the Austrian Labor Market

Ivan Pryanishnikov
Vienna University of Technology, Austria

Katarina Zigova
Vienna Institute of Demography, Austrian Academy of Sciences

Abstract: In this paper we analyze the selection of industry branches by employees in the Austrian labor market. For this purpose we use the standard logit model and the heteroscedastic extreme value model. We show that the likelihood ratio test rejects the multinomial logit model in favor of the heteroscedastic specification. Consequently, we concentrate on estimation results of the heteroscedastic extreme value model. In our investigation we use 1997 social security records provided by the Hauptverband der Sozialversicherungen.

Zusammenfassung: In dieser Arbeit analysieren wir die Wahl des Industriezweigs durch die Arbeitskräfte im österreichischen Arbeitsmarkt. Zu diesem Zweck verwenden wir das multinomiale Logit-Modell und das heteroskedastische Extremwertmodell. Mit Hilfe des likelihood Verhältnis Tests zeigen wir, dass das multinomiale Logit Modell ungeeignet ist, und untersuchen in der Folge nur die heteroskedastische Spezifikation. Für unsere Untersuchungen verwenden wir Sozialversicherungsdaten des Hauptverbands der Sozialversicherungen für das Jahr 1997.

Keywords: Labor Market, Choice Model, Unobserved Utility, Heteroscedasticity.

1 Introduction

A labor market is formed by industries which have advantages and disadvantages from the employees' point of view. The employee in turn has a set of its own characteristics. These characteristics along with its preferences about industries and its unobserved incentives are supposed to lead it to choose a particular industry. The goal of this paper is to investigate the behavior of the employee in the Austrian labor market depending on employee's properties. In particular, we address here the following issue: which properties of the decision maker (i.e., employee) change the probabilities to be employed in a certain sector. For instance, we can easily assume that being high educated increases the probability to work in a science or manufacturing sector, while living in lower developed area may increase the probability to be engaged in the agriculture. In this study, an industrial sector of the economy is a nominal dependent variable. That refers our investigation to the class of multinomial choice models.

Choice models are derived from the utility maximization hypothesis. This hypothesis assumes that a decision maker's choice is the result of its preferences. The decision maker selects the alternative with the highest preference or utility. The utility that a decision maker associates with an alternative is specified to be the sum of a deterministic and

random components. The deterministic component is a function, which depends on observed attributes of the alternative and observed individual characteristic of the decision maker. The random component is a random process representing the effect of unobserved attributes of the alternative and unobserved characteristics of the decision maker.

In most choice models, the random components of the utilities of the different alternatives are assumed to be independent and identically distributed (IID) with a type I extreme value distribution (see Johnson and Kotz, 1970). This results in the multinomial logit choice model (see McFadden, 1974). The multinomial logit model has a simple closed form structure, making it easy to estimate and interpret. However, the IID property of the multinomial logit is unlikely to represent actual choice behavior in many situations (see Stopher et al., 1981).

Inflexibility of the multinomial logit can be relaxed by removing, fully or partly, the IID assumption on the random components of the utilities of the different alternatives. The IID assumption can be relaxed in the following ways:

   (i) allowing the random components to be non-identical and non-independent,

  (ii) allowing the random components to be correlated, maintaining the assumption that they are identical distributed, and

 (iii) allowing the random components to be non-identical (different variances), but maintaining the independence assumption.

These three cases are discussed briefly.

Case (i): models with non-identical, non-independent random components are referred to the mixed logit (also called random-parameters logit) model (see McFadden and Train, 2000; Ben-Akiva et al., 2003) and the probit model (see Daganzo, 1979). The mixed logit generalizes standard logit by allowing the parameter associated with each observed variable to vary randomly across decision-makers. Variance in the unobserved decision-maker-specific parameters induces correlation over alternatives in the random component of utility. The distribution of the parameters is usually assumed to be normal, lognormal or gamma, but it can be any other. Simultaneously, that makes the mixed logit very flexible, however, on the other hand there is no economic theory which motivates the distribution selection of the unobserved decision-maker-specific parameters. Estimation of the mixed logit parameters is computationally intensive, and involves evaluation of the manifold integral with no closed form solution. The dimension of the integral is equal to the number of parameters to be estimated. The accurate Gaussian quadrature is feasible in cases of only one- or two-dimensional integration, that requires a very restrictive specification.

The multinomial probit model assumes that a normal distribution for the error terms can accommodate a very general error structure. However, the increase in flexibility of the error structure can lead to some statistical and practical difficulties, including difficulty in interpretation, non-intuitive model behavior, and low precision of covariance parameter estimates (see Horowitz, 1981). The multinomial probit choice probabilities also involve high dimensional integrals and this may pose computational problems when the number of alternatives exceeds three.

Case (*ii*): the distribution of the random components in the model which uses identical non-independent random components is usually specified as a type I extreme value distribution. The resulting model is referred to the nested logit model. This model allows partial relaxation of the independence among random components of alternatives (see Daly and Zachary, 1979; McFadden, 1978). The nested logit has closed form solution, is relatively simple to estimate. However, it requires an a priori specification of homogenous sets of alternatives for which the IID property holds. This requirement has at least two drawbacks. First, the number of different structures to estimate in a search of the best structure increases rapidly as the number of alternatives increases. Second, then actual competition structure among alternatives may be a continuum which cannot be accurately represented by partitioning the alternatives into mutually exclusive subsets.

Case (*iii*): models with independent, but not identically distributed error terms taking heteroscedasticity in alternative error terms into account are presented in the literature in various forms. Daganzo (1979) used independent negative exponential distributions with different variances for the random error components to develop a closed form discrete choice model. However, his model has not seen much application since it requires that the utility of any alternative does not exceed an upper bound. Steckel and Vanhonacker (1988) suggested a heterogeneous conditional logit model, where the error component is a mixture of type I extreme value and gamma distribution. They derived a closed form solution of the choice probability. Bhat (1995) developed a random utility model with independent, but not-identically error terms distributed with a type I extreme value distribution, allowing the utility of alternatives to differ in variances of the random components across alternatives. This model nests the multinomial logit model. Furthermore, it is flexible enough to let differ cross-elasticities among all combinations of alternatives, as the unobserved part of individual utility function (see the next section) is allowed to vary with the choices. It does not require an a priori identification of mutually exclusive partitioning as does the nested logit structure. In addition, it poses much less computationally burden, requiring only the evaluation of a 1-dimensional integral (independent of the number of alternatives) compared to the evaluation of multidimensional integral in the mixed model and multinomial probit model. Bhat (1995) applies this so-called heteroscedastic extreme value model for an intercity travel mode choice. Allenby and Ginter (1995) proposed a similar model in marketing context. However, the discussion of the model and the procedure to estimate the model are different in the two research efforts.

In this paper we apply the standard logit model McFadden (1974) and the heteroscedastic extreme value model (see Bhat, 1995; Allenby and Ginter, 1995) in the Austrian labor market choice model. The heteroscedastic extreme value model does not require any prior identification of mutually exclusive partitioning as does the nested logit structure. In addition, it poses much less computational burden, requiring only the evaluation of a 1-dimensional integral (independent of the number of alternatives) compared to the evaluation of multidimensional integral in the mixed model and multinomial probit model. The heteroscedastic extreme value model allows different variances on the random components across alternatives, that intuitively makes it more attractive than the standard logit model. Unequal variances of the random components is likely to occur when the variance of an unobserved variable is different for different alternatives (see Horowitz, 1991). For example, in a labor market choice model, if guarantee of employment is an unobserved

variable whose values vary considerably for manufacturing (based on, say, the degree of competition in different firms), but little for science, then the random component for manufacturing and science will have different variances.

The paper is organized as follows. Section 2 describes the two choice models we will apply in the paper, i.e. the multinomial logit, and the heteroscedastic extreme value model. The data used in this study is described in Section 3. Section 4 discusses the estimation results, and Section 5 concludes.

2   The Model Specification and Estimation

In this section, we describe two models used in the paper, namely, the multinomial logit, and the heteroscedastic extreme value model. We overview the utility maximization hypothesis on which these models are based, and sketch the estimation procedure of the models. As well as a method of the model interpretation is described.

2.1   The Random Utility Model

Let decision-maker n choose from a set of mutually exclusive alternatives, $j = 1, ..., J$. The decision-maker obtains a certain level of utility U_{nj} from each alternative. The discrete choice model is based on the principle that the decision-maker chooses the outcome that maximizes the utility. We do not observe her utility, but observe some attributes of the alternatives as faced by the decision-maker. Hence, the utility is decomposed into deterministic V_{nj} and random part ε_{nj}:

$$U_{nj} = V_{nj} + \varepsilon_{nj}, \ \forall j. \tag{1}$$

Since ε_{nj} is not observed, the decision-maker's choice cannot be predicted exactly. Instead, the probability of any particular outcome is derived. The unobserved term is treated as random with density $f(\varepsilon_{nj})$. The joint density of the random vector $\varepsilon_n = \langle \varepsilon_{n1}, ..., \varepsilon_{nJ} \rangle$ is denoted $f(\varepsilon_n)$. Probability that decision-maker n chooses alternative i among J alternatives is

$$
\begin{aligned}
P_{ni} &= \Pr(U_{ni} > U_{nj} \ \forall j \neq i) \\
&= \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \ \forall j \neq i) \\
&= \int I(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \ \forall j \neq i) f(\varepsilon_n) d\varepsilon_n,
\end{aligned}
\tag{2}
$$

where $I(\cdot)$ is the indicator function, equaling 1 when the term in parenthesis is true and 0 otherwise. This is a multidimensional integral over the density of the unobserved portion of utility $f(\varepsilon_n)$. Different discrete choice models are obtained from different specifications of the density. The deterministic part V_{nj} of utility is usually treated as a linear function of explanatory variables x and an unknown vector of underlying parameters θ. In random utility models the expectation of the random component $E(\varepsilon_{nj})$ is assumed to equal 0, that in turn implies $E(U_{nj}) = V_{nj}$. A vector of utilities $U_{nj}, \forall j$ is assumed to be continuously distributed with an existing covariance matrix (see Tutz, 2000).

The absolute level of utility in Equation 2 is irrelevant to the decision maker behavior. For example, if a constant is added to the utility of all alternatives, the alternative with the highest utility does not change. The choice probability is $P_{ni} = P(U_{ni} > U_{nj}) = P(U_{ni} - U_{nj} > 0)$, which depends only on the difference in utility, not its absolute level. The fact that only differences in utility matter has several implications for the identification and specification of discrete choice models. In general it means that the only parameters that can be estimated (that is, are identified) are those that capture differences across alternatives.

In order to investigate the way how observed factors influence the decision maker to make a choice, unknown parameters θ of the model are estimated. The log-likelihood estimator can be used to estimate the parameters. The log-likelihood function to be maximized over parameters θ is given:

$$\ln L(\theta) = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \ln P_{nj}. \tag{3}$$

Where y_{nj} equals 1 if alternative j is chosen and equals 0 for all other non-chosen alternatives.

2.2 The Multinomial Logit

The multinomial logit (MNL) model, invented by McFadden (1974), is obtained by the assumption that each random components ε_{nj} in the utilities (1) is distributed independently, identically type I extreme value, where the variance of the error term is equal to $\pi^2/6$. The density for each unobserved component of utility and the cumulative distribution are given, respectively, by

$$\lambda(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \quad \text{and} \quad \Lambda(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}. \tag{4}$$

The random utility (1) is combined with the probability distribution for the random components ε_{nj} in equation (4) and assume independence among the random components of the different alternatives. The probability that an decision maker n chooses alternative i among the J alternatives is given by

$$P_{ni} = \Pr(\varepsilon_{nj} < V_{ni} - V_{nj} + \varepsilon_{ni} \; \forall j \neq i)$$
$$= \int \prod_{j \neq i} \Lambda \left(V_{ni} - V_{nj} + \varepsilon_{ni} \right) \lambda \left(\varepsilon_{ni} \right) d\varepsilon_{ni}. \tag{5}$$

Thus, the choice probability is the integral over all values of ε_{ni} weighted by its density $\lambda(\cdot)$ as defined in (4). This integral has a closed form solution and after some manipulation the logit probabilities, with $V_{nj} = x'_n \beta_j$, become:

$$P_{ni} = \frac{e^{x'_n \beta_i}}{\sum_j e^{x'_n \beta_j}}. \tag{6}$$

Since MNL is a model where regressors do not vary over choices, coefficients are estimated for any choice. MNL requires identification: one of the choices, say j, is treated as

the base category (correspondent β_j is constrained to equal 0). Substitution equation (6) into (3) yields the log-likelihood function to be maximized over parameters β

$$\ln L(\beta) = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \ln \frac{e^{x'_n \beta_i}}{\sum_j e^{x'_n \beta_j}}. \tag{7}$$

Estimation results of the multinomial logit model for the Austrian labor market is given in Table A.1 in the appendix. The base category is *manufacturing*.

2.3 The Heteroscedastic Model

The heteroscedastic model is derived under the assumption that the random components in the utilities (1) are independent, but not-identically distributed. The random components is assumed to have a location parameter equal to 0 and a scale parameter equal to γ_j for the jth alternative, then the variance of the jth alternative's error term is $\pi^2 \gamma_j^2/6$. The assumption of a location parameter equal to zero for the random components is not restrictive since constants are included in the systematic utility for each alternative. Thus, the probability density function $f(\cdot)$ and the cumulative distribution function $F(\cdot)$ of the random error for the jth alternative are, respectively,

$$f(\varepsilon_{nj}) = \frac{1}{\gamma_j} e^{-\frac{\varepsilon_{nj}}{\gamma_j}} e^{-e^{-\varepsilon_{nj}}} \ \text{ and } \ F(\varepsilon_{nj}) = e^{-e^{-\frac{\varepsilon_{nj}}{\gamma_j}}}. \tag{8}$$

The random utility (1) combined with the assumed probability distribution for the random components in equation (8) and assumed independence among the random components of the different alternatives, enables to develop the probability that a decision maker n chooses alternative i form the J alternatives

$$\begin{aligned}
P_{ni} &= \Pr(\varepsilon_{nj} < V_{ni} - V_{nj} + \varepsilon_{ni} \ \forall j \neq i) \\
&= \int \prod_{j \neq i} \Lambda\left(\frac{V_{ni} - V_{nj} + \varepsilon_{ni}}{\gamma_i}\right) \frac{1}{\gamma_i} \lambda\left(\frac{\varepsilon_{ni}}{\gamma_i}\right) d\varepsilon_{ni} \\
&= \int \prod_{j \neq i} \Lambda\left(\frac{V_{ni} - V_{nj} + \gamma_i w_n}{\gamma_i}\right) \frac{1}{\gamma_i} \lambda(w_n) dw_n
\end{aligned} \tag{9}$$

where $\Lambda(\cdot)$ and $\lambda(\cdot)$ are given by equation (4), and $w_n = \varepsilon_{ni}/\gamma_i$. The probabilities given by equation (9) sum up to one over all alternatives (Bhat, 1995). If the scale parameters of the random components of all alternatives are equal, then the probability expression in equation (9) collapses to the multinomial logit (McFadden, 1974).

The heteroscedastic model is estimated using the maximum likelihood technique. As before $V_{nj} = x'_n \beta_j$. The parameters to be estimated in the model are the parameter vector β and the scale parameters γ of the random component of each alternative (one of the scale parameters is normalized to one for identification). The log-likelihood function to be maximized is given as

$$\ln L(\beta, \gamma) = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \ln \int \prod_{j \neq i} \Lambda\left(\frac{V_{ni} - V_{nj} + \gamma_i w_n}{\gamma_i}\right) \frac{1}{\gamma_i} \lambda(w_n) dw_n. \tag{10}$$

The log-likelihood function (10) has no closed form solution. The integral has to be computed for each alternative-decision maker combination at each iteration of the maximization of the log-likelihood function.

To maximize the log-likelihood function, we use the Newton-Rhapson maximization. The idea of the method is to find the value β and γ that maximize $\ln L(\beta, \gamma)$. Numerically, the maximum can be found by "walking up" the likelihood function until no further increase can be found. Each iteration moves to a new value of the parameters at which $\ln L(\beta, \gamma)$ is higher than at the previous value. A new value of coefficients β_{t+1} and γ_{t+1} is given by $(-H_t^{-1})g_t$, where g_t and H_t are, respectively, gradient (i.e., the vector of first derivatives) and Hessian (i.e, the matrix of second derivatives) of $\ln L(\beta_t, \gamma_t)$ evaluated at β_t, γ_t. For further details see, for example, Judge et al. (1980).

The estimation of the log-likelihood function involves a one-dimension integral with no close form solution. The Gaussian quadrature is used to obtain an accurate approximation of the integral in (10). The idea of the Gaussian quadrature is based on the device of adding up the value of the integrand at a sequence of abscissas within the range of integration. To evaluate an integral of a function $f(x)$ the following approximation is used:

$$\int_a^b f(x)dx = \int_a^b p(x)w(x)dx \approx \sum_{i=1}^n w_i p(x_i) \tag{11}$$

where $p(x)$ is polynomial of degree $2n - 1$ or lower, and $w(x)$ is a chosen basis function, or weight. The computation of the integral involves two distinct phases: (i) the generation of the orthogonal polynomials $p(x)$ i.e., the computation of the coefficients; (ii) the determination of the zeros of $p(x)$, and the computation of the associated weights. The advantage of the Gaussian quadrature is the freedom to choose not only the weighting coefficients, but also the location of the abscissas at which the function is to be evaluated. For an extensive overview on the numerical integration, the reader is referred to Geweke (1996), or Press et al. (1992). In our analysis, the numerical evaluation of the integral was done with an in-built Gauss procedure.

Estimation results of the heteroscedastic extreme value logit for the Austrian labor market choice model is given in Table A.1 in the appendix. As for MNL the based category is *manufacturing*.

2.4 Interpretation of Parameters

The amount of parameters in the multinomial logit model as well as in the heterogenous extreme value model increases with the number of outcomes and the number of independent variables and hence it is usually very large. Magnitudes and signs of parameters are hardly directly informative.

In this paper interpretation of parameters of the model is based on a discrete change in predicted probabilities (see Scott Long, 1997). The probability that the decision maker chooses alternative i from J alternatives is given by equations (5) for the multinomial logit and (9) for the heteroscedastic model, in which substituting $\widehat{\beta}$ instead of β yields the predicted probability. The discrete change in the predicted probability occurs when an explanatory variable, say x_k, changes from x_s (for the starting value) to x_e (for the ending

value):

$$\frac{\triangle \Pr(y = i|x)}{\triangle x} = \Pr(y = i|x_e) - \Pr(y = i|x_s).$$

The predicted probabilities on the right hand side of the expression are calculated holding all other variables, except x_k, constant.

3 Data Design

To develop a labor market choice model we use a sample from the social security records in Austria (Hauptverband der Sozialversicherungen, 1997). This records cover individual characteristics of employees (decision-makers) such as age, gender, place of residence, field of action, wage, number of employed days a year, etc, from 1984 to 1998. The sample used in this study observes 3234 employees in 1997.

The observed dependent variable is the industry which the decision-maker chooses to work in (or employers of the decision-makers). The six industry categories are: *(1) agriculture, (2) manufacturing, (3) service, (4) science, (5) public administration, (6) public health*. These six industries form a set of mutually exclusive and exhaustive alternatives from employees' point of view.

Table 1: Descriptive statistics of the data (3234 observations)

Explanatory Variable	Mean	St. Dev.	Min	Max
Age/100	0.38	0.10	0.17	0.71
Gender	64 %		0.00	1.00
High-Education Indicator	22 %		0.00	1.00
Middle-Education Indicator	59 %		0.00	1.00
High-Developed Land Indicator	49 %		0.00	1.00
Middle-Developed Land Indicator	15 %		0.00	1.00

To explain a choice of the decision-maker, six explanatory variables are used: *age, gender*, and four dummy variables. The first two control education: *high-education indicator* and *middle-education indicator*, the other two dummies *high-developed land indicator* and *middle-developed land indicator* are regional factors of development where the decision-maker lives. For convenience, variable *age* is premultiplied by 10^{-2}. In dummy variable *gender*, 1 is reserved for men. The data available do not contain a direct control variable for education, instead two dummies are constructed using ratio of two explanatory variables, *wage* over *age*. The largest 716 elements of the ratio are assigned 1 in *high-education indicator* dummy, the next 1319 largest elements of the ratio are assigned 1 in *middle-education indicator* dummy and for the remaining elements both dummies are set to 0. The dummy *high-developed land indicator* equals 1 if the decision-maker lives in the most developed region of Austria, and equals 0 otherwise. The dummy *middle-developed land indicator* equals 1 if the decision-maker lives in the middle developed region of Austria, and equals 0 otherwise. The words "most developed" and "middle developed" reflect a level of the regional GDP per capita.

Table 1 summarizes the variables used in the model with descriptive statistics. For dummy variables only the percentage of the decision-makers with value 1 is indicated, whereas for the continuous variables *age* the standard deviation is also given. To specify for a set of explanatory variables, we have performed several tests investigating the mixed effect of age, gender, and education. Since the mixed specifications were rejected, the final set of variables contains only separate variables.

4   Estimation Results and Discussion

We match the decision-maker behavior in the Austrian labor market for the year 1997. The decision-maker chooses an industry among a set of the mutually exclusive choices described above. We computed two models: the multinomial logit model and the heteroscedastic extreme value model. The parameter estimates in the first model were obtained with an own maximum likelihood routine programmed in Gauss. We verified these results with the ones obtained by Stata package. The codes underlying to the latter model were also performed in Gauss. As mentioned before, we maximized the likelihood function (10) with the own programmed Newton-Rhapson method, while the inner integral was computed with an in-built Gauss procedure.

The final estimation results (parameters, standard deviations, and P-value) are shown in Table A.1 in the appendix for both the multinomial model and the heteroscedastic extreme value model. Moreover, in Figure A.1 we graphically compare the parameters arose from the both specifications. The asymptotic covariance matrix of parameters in both estimations computed as $H^{-1}BH^{-1}$, where H is the hessian and B is the cross-product matrix of the gradients, provides consistent standard errors (see Börsch-Supan, 1987).

A comparison of the multinomial logit and the heteroscedastic extreme value model is based on the likelihood ratio test (see Greene, 2000). The multinomial logit is rejected in favor of the heteroscedastic specification. Here, the null hypothesis is defined as equality of the random terms ε_{nj} in the utility function (1) across the sectors, i.e. $H_0 : \varepsilon_{nj} = \varepsilon_{ni}$ for all $j \neq i$. This is equivalent with setting the $\gamma_j = 1$ (see in 8) for all j's. The test statistics is 12.23 which is significant at any reasonable level of significance when compared to a chi-squared statistics with five degrees of freedom. The rejection of the multinomial logit confirms the assumption about unequal variances of the random components made earlier. Hence, in the subsequent discussion we will concentrate on interpretation of the model parameters of the heteroscedastic extreme value model.

Three tests, based on the Wald methodology (see Scott Long, 1997), requiring only estimation of a single model, are run to test the parameters of the model. The null hypothesis of the first test is that a variable has no effect throughout the alternatives, i.e. $H_0 : \beta_{n\cdot} = 0$, where $\beta_{n\cdot}$ is the parameter vector underlying to the variable n. The hypothesis is rejected for all of the variables (Table 2). The null hypothesis of the second test is whether all parameters for a choice j equal to 0, i.e. $H_0 : \beta_{\cdot j} = 0$, where $\beta_{\cdot j}$ is the parameter vector underlying to the choice j. This assumption is also rejected for all choices (Table 3). The last is a test of whether a pair of choices is indistinguishable with respect to the variables and, hence, can be combined. The underlying null hypothesis is

formulated as $H_0 : \beta_{.j} = \beta_{.i}$ for $j \neq i$, where $\beta_{.j}$ is as in the second test. This hypothesis is easily rejected for all possible pairs (Table 4).

Table 2: Chi square values of the test that a variable has no effect throughout the alternatives

	Age	Gender	High-Education Indicator	Middle-Education Indicator	High-Develop. Land Ind.	Middle-Develop. Land Ind.
$\chi^2(5)$	24.43	229.44	38.73	21.12	109.52	19.23
P-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 3: Chi square values of the test whether all parameters for a choice are equal to 0

	Agriculture	Service	Science	Public Adm.	Public Health
$\chi^2(6)$	18.66	160.93	116.18	172.89	169.55
P-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 4: Chi square values of the test that a pair of choices is indistinguishable with respect to the variables

		Service	Science	Public Adm.	Public Health
Agriculture	$\chi^2(6)$	21.13	27.31	28.15	41.02
	P-value	(0.00)	(0.00)	(0.00)	(0.00)
Service	$\chi^2(6)$		19.92	36.46	96.16
	P-value		(0.00)	(0.00)	(0.00)
Science	$\chi^2(6)$			26.83	76.62
	P-value			(0.00)	(0.00)
Public Adm.	$\chi^2(6)$				64.53
	P-value				(0.00)

To examine the effects of the explanatory variables of the model, we use the difference in predicted probabilities (see Subsection 2.4). Table 5 contains estimates of discrete change in predicted probabilities from the heteroscedastic extreme value model.

First, consider the dummy variable *gender* (recall that 1 is reserved for men). The effect of *gender* is largest on the probabilities of working in *manufacturing* (0.30). On the contrary being woman increases the probability to be engaged in all the other industries except *agriculture*, with the largest effect on the probability to work in *public health* (0.14).

Next, let us consider the dummy variable *high-education indicator*. Holding all other variables constant, being high educated decreases the probability to work in *service* by 0.11, in *public administration* by 0.08, in *agriculture* by 0.06, and increases the probability to be engaged in *manufacturing* by 0.14, in *science* by 0.07, in *public health* by

0.04. The *Middle-education indicator* demonstrates almost the same tendency as the *high-education indicator* with weaker changes in probabilities. The *Middle-education indicator* positively changes only the sign in the sector *public administration* indicating the inflow of middle educated employees in the *public administration* sector.

Now, we turn our attention on how the industrial development of the area where the decision-maker lives influences its choice. Living in high-developed land decreases the probability to work in *manufacturing* by 0.22. In comparison *service*, *science* and *public administration* increase the probabilities to choose by 0.17, 0.09, 0.09, respectively. Only change in the probability to work in *public health* changes from increasing (for *high-developed land indicator*) to decreasing (for *middle-developed land indicator*).

Table 5: Discrete changes in predicted probabilities of the heterogenous extreme value model

	Age		Gender	High-Education Indicator	Middle-Education Indicator	High-Develop. Land Ind.	Middle-Develop. Land Ind.
	$25 \to 40 \to 55$		$0 \to 1$	$0 \to 1$	$0 \to 1$	$0 \to 1$	$0 \to 1$
Agriculture	-0.04	-0.02	0.05	-0.06	-0.03	-0.05	0.02
Manufacturing	0.06	0.07	0.30	0.14	0.07	-0.22	-0.16
Service	-0.07	-0.09	-0.06	-0.11	-0.06	0.17	0.06
Science	-0.02	-0.04	-0.05	0.07	-0.04	0.09	0.03
Public Adm.	0.03	0.04	-0.10	-0.08	0.03	0.09	0.03
Public Health	0.04	0.04	-0.14	0.04	0.03	-0.08	0.02

The model contains only one continuous variable: *age*. To examine the effect of *age* we consider two discrete changes, first, from 25 to 40 years, second, from 40 to 55 years. We see that the tendency in the both columns in Table 5 is the same for the 6 sectors except *agriculture*, where the sign of the discrete change in the probability remains the same, and the effect softens up. On the contrary, the effect for all the remaining alternatives strengthens with ages. Getting older in the age class from 40 to 55 increases the probability to work in *manufacturing* by 0.07, in *public administration* by 0.04, in *public health* by 0.04, and decreases the probability to work in the *service* sector by 0.09, in *science* by 0.04 and in *agriculture* by 0.02. The largest effect of *age* is observed in the *manufacturing* (increasing) and in the *service* sector (decreasing).

To a large degree the effect of the variables in the model corresponds to economic intuition. For instance, it is natural to suppose, that being man increases the probability to work in *manufacturing*, and being woman increases the probability to work in *public health*. Also it is sensible to expect that getting older only strengthens any effect, keeping the sign unaffected. However, the estimation results reveal some effects that are rather surprising. One would expect, e.g. that living in the low-developed area decreases the probability to choose the manufacturing sector, while in our analysis the reverse holds. The probability to choose the manufacturing sector increases by 0.38.

5   Conclusion

In this paper we applied the multinomial logit model and the heterogenous extreme value model to the problem of employee's choice of her preferred industry. We grouped all employers under six industries, namely agriculture, manufacturing, service, science, public administration and public health. As explanatory variable we chose 4 factors, i.e. age, gender, education and economic development of her residence, representing the characteristics of the employee.

Using a likelihood ratio test, we found that the heterogenous extreme value model is more appropriate for the data. This conclusion is not surprising as this model allows for extra variation in the random components of the utility function (1) across the industrial sectors. The differences in predicted probabilities support also some plausible decision making. For instance, being high educated increases the probability to work in manufacturing about 0.14, while the increase of probability to work in the scientific sector is rather low (0.07). We found some unexpected results, e.g. that living in the low-developed area increases the probability to choose the manufacturing sector by 0.38. This might be because of lack of other alternatives.

In the paper we demonstrated the usefulness of the multiple choice model for modelling employee's job decisions. As a further extension one could apply another choice model, the mixed logit model (see McFadden and Train, 2000; Ben-Akiva et al., 2003), which allows in addition, for non-independent random components of the utility function.

A   Appendix

Table A.1: Estimation results of the multinomial logit and heteroscedastic extreme value model (manufacturing is base)

| Variable	Multinomial Logit			Heteroscedastic Extreme Value Model		
	Parameter	Std.Err.	P-value	Parameter	Std.Err.	P-value
Constant						
Agriculture	-2.5024	1.0479	0.01	-1.7822	0.6348	0.00
Service	1.0274	0.2381	0.00	1.8126	0.2096	0.00
Science	-0.6595	0.2811	0.00	-1.5219	0.5421	0.00
Public Administration	-0.6145	0.2451	0.00	-1.1261	0.4252	0.00
Public Health	-1.0454	0.4652	0.02	-0.4521	0.2144	0.00
Age						
Agriculture	-3.1320	1.4291	0.05	-3.3209	1.5089	0.01
Service	-1.9124	0.6322	0.00	-1.7199	0.4671	0.00
Science	-1.6348	0.6342	0.00	-1.3942	0.6505	0.02
Public Administration	-0.2409	0.1103	0.03	-0.3309	0.2012	0.10
Public Health	1.5112	0.9121	0.09	1.1105	0.5114	0.03
Gender						
Agriculture	1.1038	0.7857	0.16	0.9611	0.5114	0.06
Service	-0.4921	0.0613	0.00	-0.8900	0.0964	0.00
Science	-1.3490	0.1265	0.00	-1.0490	0.1445	0.00
Public Administration	-1.1128	0.1172	0.00	-1.3193	0.1345	0.00
Public Health	-2.7601	0.3541	0.00	-2.3032	0.1874	0.00
High-Educ.Indicator						
Agriculture	-2.6881	0.9952	0.00	-3.1961	1.1190	0.00
Service	-0.3441	0.1056	0.00	-0.5550	0.1556	0.00
Science	-0.1203	0.0471	0.00	-0.1591	0.0683	0.00
Public Administration	-0.6301	0.2932	0.03	-0.7939	0.2421	0.00
Public Health	0.1428	0.0893	0.11	0.1083	0.0591	0.06
Middle-Educ.Indicator						
Agriculture	-0.4431	0.2546	0.08	-0.6973	0.3892	0.07
Service	-0.6217	0.2703	0.02	-0.2787	0.1042	0.00
Science	-0.5428	0.3568	0.12	-0.2839	0.1941	0.06
Public Administration	-0.3721	0.2054	0.07	-0.1164	0.1718	0.11
Public Health	-0.2329	0.1502	0.12	-0.1074	0.0551	0.05
High-Develop.Land Indic.						
Agriculture	0.3219	0.6347	0.04	0.1721	0.4266	0.01
Service	0.5201	0.0346	0.00	0.6851	0.0941	0.00
Science	1.4214	0.2368	0.00	1.0649	0.1589	0.00
Public Administration	1.1621	0.1931	0.00	0.9022	0.1449	0.00
Public Health	-0.3525	0.1356	0.00	-0.3052	0.1279	0.00
Middle-Develop.Land Indic.						
Agriculture	0.7121	0.7178	0.32	0.5571	0.4855	0.25
Service	0.1253	0.0539	0.00	0.2980	0.1156	0.00
Science	0.3726	0.1038	0.00	0.4444	0.1827	0.00
Public Administration	0.4112	0.2648	0.00	0.3648	0.1737	0.00
Public Health	0.4515	0.1430	0.03	0.3902	0.1537	0.04
Scale Parameter						
Agriculture	1.0000			2.5354	1.0986	0.02
Service	1.0000			0.8324	0.2017	0.00
Science	1.0000			1.2493	0.4302	0.00
Public Administration	1.0000			0.7543	0.3002	0.00
Public Health	1.0000			1.6850	0.8245	0.04
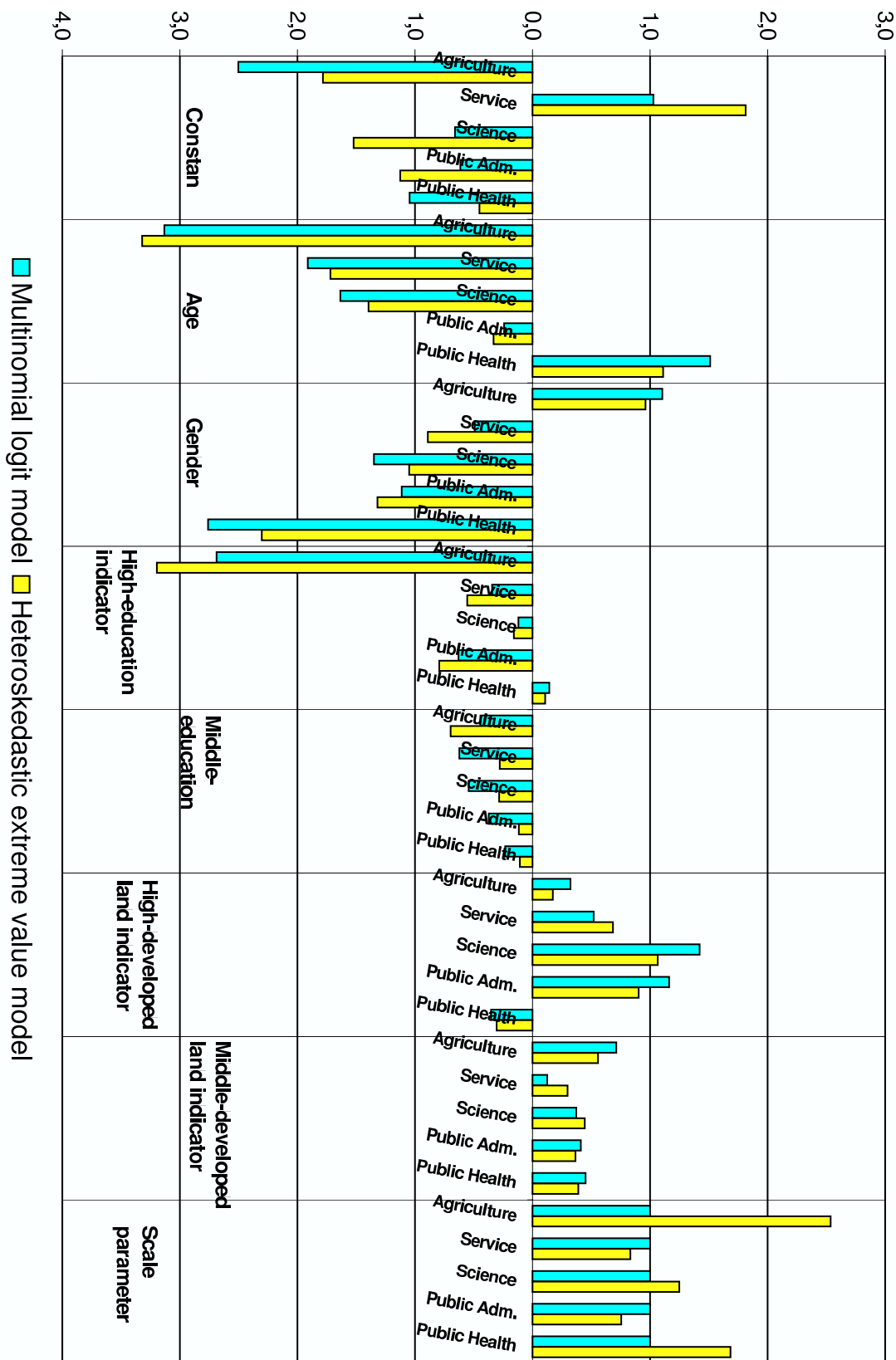
Figure A.1: Graphical comparison of the parameter estimates for the Multinomial Logit Model and the Heterogenous Extreme Value Models

References

G. Allenby and J. Ginter. Modelling competitive subsets and product differentiation. *International Journal of Research in Marketing*, 12, 1995.

M. Ben-Akiva, D. Bolduc, and J. Walker. Identification of the logit kernel (or mixed logit) model. *Working Paper, Department of Civil and Environmental Engineering, MIT*, 2003.

C. Bhat. A heteroscedastic extreme value model of intercity choice. *Transportation Research B*, 29, 1995.

A. Börsch-Supan. *Econometric Analysis of Discrete Choice. With Application on the Demand for Housing in the United States and West Germany. Lecture Notes in Economics and Mathematical Systems 296.* Springer, Heidelberg, New York, Tokyo, 1987.

C. Daganzo. *Multinomial Probit: The Theory and Its Application to Demand Forecasting.* Academic Press, New York, 1979.

A. Daly and S. Zachary. Improved multiple choice models. In D. Hensher and Q. Davi, editors, *Identifying and Measuring the Determinants of Mode Choice*. Teakfield, London, 1979.

J. Geweke. Monte carlo simulation and numerical integration. In H. Amman, D. Kendrick, and J. Rust, editors, *Handbook of Computational Economics*, pages 731–800. Elsevier Science, Amsterdam, 1996.

W. Greene. *Econometric Analysis*. Prentice Hall, Upper Saddle River, New Jersey, 4th edition, 2000.

J. Horowitz. Identification and diagnosis of specification errors in the multinomial logit model. *Transportation Research B*, 15, 1981.

J. Horowitz. Reconsidering the multinomial probit model. *Transportation Research B*, 25, 1991.

N. Johnson and S. Kotz. *Distributions in Statistics: Continuous Univariate Distributions*. John Wiley & Sons, New York, 1970.

G. Judge, W. Griffiths, R. Hill, and T. Lee. *The Theory and Practice of Econometrics*. John Wiley & Sons, New York, 1980.

D. McFadden. Conditional logit analysis of qualitative choice behaviour. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974.

D. McFadden. Modelling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, editors, *Spatial Interaction Theory and Planning Models*, pages 75–96. North Holland, Amsterdam, 1978.

D. McFadden and K. Train. Mixed mnl models of discrete response. *Journal of Applied Econometrics*, 15(5):447–470, 2000.

W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press, Cambridge, New York, 2nd edition, 1992.

J. Scott Long. *Regression Models for Categorical and Limited Dependent Variables*. SAGE Publications, Volume 7, 1997.

J. Steckel and W. Vanhonacker. A heterogenous conditional logit model of choice. *Journal of Business and Economic Statistics*, 6, 1988.

P. Stopher, A. Meyburg, and W. Brög. Travel behaviour research: a perspective. In P. Stopher, A. Meyburg, and W. Brög, editors, *New Horizons in Travel Behaviour Research*. Lexington Books, Lexington, 1981.

G. Tutz. *Die Analyse kategorialer Daten: anwendungsorientierte Einführung in Logit-Modellierung und kategoriale Regression*. Oldenbourg Wissenschaftsverlag GmbH, München, 2000.

Authors' addresses:

Ivan Pryanishnikov   Katarina Zigova
Department of Computer Languages   Vienna Institute of Demography
Vienna University of Technology   Austrian Academy of Sciences
Argentinierstr. 8   Prinz-Eugen Str. 8
A-1040 Vienna   A-1040 Vienna
Austria   Austria

Tel. +43 1 58801 / 58520   Tel. +43 1 51581 / 7712
Fax +43 1 58801 / 18598   Fax +43 1 51581 / 7730
E-mail: prianich@complang.tuwien.ac.at   E-mail: katarina.zigova@assoc.oeaw.ac.at