Australian School of Business Research Paper No. 2012 ECON 46

# The perceived unreliability of rank-ordered data: an econometric origin and implications

Hong il Yoo

# The perceived unreliability of rank-ordered data: an econometric origin and implications

Hong il Yoo*

(h.yoo@unsw.edu.au)

School of Economics, University of New South Wales

November 4, 2012

**Abstract**

The problem of unstable coefficients in the rank-ordered logit model has been traditionally interpreted as a sign that survey respondents fail to provide reliable ranking responses. This paper shows that the problem may embody the inherent sensitivity of the model to stochastic misspecification instead. Even a minor departure from the postulated random utility function can induce the problem, for instance when rank-ordered logit is estimated whereas the true additive disturbance is iid normal over alternatives. Related implications for substantive analyses and further modelling are explored. In general, a well-specified random coefficient rank-ordered logit model can mitigate, though not eliminate, the problem and produce analytically useful results. The model can also be generalised to be more suitable for forecasting purposes, by accommodating that stochastic misspecification matters less for individuals with more deterministic preferences. An empirical analysis using an Australian nursing job preferences survey shows that the estimates behave in accordance with these implications.

**JEL classification:** C25, C52, C81

**Key words:** rank-ordered logit, mixed logit, latent class, stated ranking

*School of Economics, Level 4, Australian School of Business Bldg., UNSW, Sydney NSW 2052, Australia. email: h.yoo@unsw.edu.au. Ph: +61 2 9385 2660. Fax: +61 2 9313 6337.

# 1 Introduction

The use of stated preference surveys has become commonplace in the discrete choice modelling literature, as demonstrated by the lists of cited applications in popular econometrics textbooks (Greene, 2008; Train, 2009). These surveys provide, often the only, practical means to collect data for analysing consumer preferences for non-market goods and potential market goods which are yet to be introduced (Vossler *et al.*, 2012). Stated preference data have extended the range of questions which can be empirically addressed in areas characterised by the scarcity of adequate revealed preference data, including environmental economics, health economics and transportation economics.

A rank-ordered dependent variable indicates a ranking of different objects from best to worst, and is more commonly, though not only, encountered in stated preference analysis than in non-experimental contexts.[1] A scenario in a typical stated preference survey comprises a small number of hypothetical alternatives with differentiated characteristics. Individual respondents are usually prompted to choose one, or state the best, alternative from each presented scenario, but can also be asked to rank all alternatives in the given scenario simply by modifying the prompt question. Econometric models for rank-ordered data can be derived from the same random utility models as their counterparts for multinomial choice data, the most popular among them being the rank-ordered logit (ROL) model (Beggs *et al.*, 1981) that builds on the multinomial logit (MNL) model (McFadden, 1981). Rankings provide more information than choices in regard to how variations in the observed characteristics influence preferences over the alternatives, allowing more precise estimation of the underlying utility coefficients.

A long standing issue in rank-ordered data analysis has been what Foster and Mourato (2002) term the problem of unstable coefficients across ranks in ROL. For example, suppose that full rankings of 4 alternatives have been observed. The data can then be recoded as though only the best alternative has been observed, or the best and the second-best have been observed. The coefficients of a correctly specified ROL model can be consistently estimated using any of the recoded and original responses. As presented in several empirical studies since Chapman and Staelin (1982), however, the ROL estimates tend to vary substantially following such recoding. In particular, the estimates typically become attenuated monotonically as each worse-ranked alter-

[1]See Berry *et al.* (2004) and Train and Winston (2007) for examples of non-experimental rank-ordered data.

native is successively incorporated into the dependent variable, as if such information increases the residual variance (Hausman and Ruud, 1987).

Over years, the problem of unstable coefficients has been explained as resulting from intra-personal heterogeneity in respondent behaviour. Hausman and Ruud (1987) formulate the heteroskedastic ROL (HROL) model, hypothesising that respondents state better-ranked alternatives with more certainty. Ben-Akiva *et al.* (1992) speculate that respondents adopt different decision protocols when stating better- and worse-ranked alternatives, e.g. due to justification bias, and generalise HROL accordingly. Fok *et al.* (2012) specify a latent class ROL model assuming that some respondents rank inferior alternatives arbitrarily due to the lack of capabilities to discriminate among them. In a similar vein, many studies have exploited specialised survey designs to investigate the relative reliability of information on better- and worst-ranked alternatives, and that of rank-ordered and choice data (see Boyle *et al.*, 2001; Foster and Mourato, 2002; Caparros *et al.*, 2008; Scarpa *et al.*, 2011; and references therein).

This paper is motivated by the rarely acknowledged distinction between the informational contents of rank-ordered data assumed by Beggs *et al.* (1981) and studies addressing the problem of unstable coefficients. The former follows the microeconomic approach to discrete choice analysis (McFadden, 1981; Anderson *et al.*, 1992). An individual's preference relation on alternatives is *ex-ante* random, and her rank-ordered response is a realised preference relation. The data generating process for all ranks of her response is one and the same. The latter studies typically cast a rank-ordered response as a sequence of independent choice outcomes, constructed in a similar manner as the top-down model of ranking behaviour from psychology (pp.69-70, Luce, 1959). An individual probabilistically chooses the best of all alternatives, excludes it from further consideration, and then chooses the best among the remaining alternatives to identify her second-best, and so on. The data generating process for each rank of her response may vary, reflecting intra-personal behavioural heterogeneity.

From the microeconomic perspective, intra-personal heterogeneity in respondent behaviour is thus an unlikely origin of unstable coefficients across ranks, and the associated explanations may be questioned on two grounds. First, by itself, the problem of of unstable coefficients is simply evidence against the postulated logit utility specification; it is not a symptom of behavioural response patterns compromising 'consistency' (Foster and Mourato, 2002) and 'reliability' (Ben-Akiva *et al.*, 1992) of rank-ordered data. Second, invoking the inadequate modelling of a response construction sequence as specification error amounts to invalidating other models for rank-ordered data derived from the microeconomic approach, including rank-ordered probit (p.158, Train, 2009) and nested ROL (Daksvik and Liu, 2009) models; unlike ROL, these models cannot be motivated and therefore extended as a probabilistic model of a sequentially constructed ranking. Such conclusion is undue when the motivating empirical regularity concerns the ROL model alone.

This paper offers an explanation for the problem of unstable coefficients from the microeconomic perspective, by exploring its origin within the random utility function motivating the ROL model. Our simulated evidence exposes an empirically relevant facet of the ROL model which, to the best of our knowledge, has not been explicitly discussed in the literature. Specifically, a slight departure from the postulated random utility distribution is enough to induce the coefficient estimates to become unstable across ranks. This includes the case when the true additive error is independently and identically distributed (iid) normal over alternatives, even though the difference between independent normal and extreme value errors are often empirically indistinguishable when modelling other types of limited dependent variables. One practical implication is that unstable coefficients can be expected in almost any empirical work, regardless of data reliability, when ROL is employed as a tractable approximation to an unknown data generating process.

The sensitivity of the ROL model to stochastic misspecification arises from the fact that the ranking probability becomes a product of multinomial choice probabilities when the error terms are iid extreme value. This product structure embodies the independence-of-irrelevant-alternatives (IIA) property that is not shared by ranking probabilities derived from most other distributions (p.157, Train, 2009). In consequence, a misspecified ROL model does not mimic the behaviour of a true ranking probability model directly. Instead, it attempts to mimic that of several choice probability models using a single set of coefficients, whereas each of these models may have different coefficients. Because a different subset of the choice probability models is mimicked depending on how the dependent variable is recoded, the resulting sets of ROL estimates would look as though respondents stated better- and worse-ranked alternatives in a different manner.

Our findings analytically complement Layton's (2000) empirically motivated conjecture that specifying the random coefficient or mixed ROL model would mitigate the problem of unstable coefficients. Stochastic misspecification can result from various sources, including unobserved preference heterogeneity. Unless such heterogeneity is exactly incorporated and no other specification error is present, the mixed ROL model would produce unstable estimates across ranks because it still assumes that the residual parts of utility are iid extreme value. Nonetheless, it can mitigate the problem because providing a good approximation to unobserved heterogeneity increases the modelled parts of utility relative to the residual parts, so that stochastic misspecification matters less in predicting preference relations.

One mainly novel modelling implication follows from the likely sensitivity of the mixed ROL to stochastic misspecification. When post-estimation analysis involving choice probabilities is a major concern, the researcher may consider estimating a flexible mixed HROL model, which allows for random heterogeneity in both coefficients and rank-specific scale parameters. Heterogeneity in both dimensions are needed because some coefficient configurations or preferences imply more deterministic individual behaviour than others; if the rank-specific scales account for the consequences of misspecifying the residuals as our analysis suggests, they should vary less across ranks for individuals with more deterministic preferences.

We also investigate whether estimates from a real application behave in accordance with the view that unstable coefficients result from stochastic misspecification. The empirical analysis uses stated preference data collected as part of an ongoing longitudinal survey of nursing students and graduates in Australia (Kenny *et al.*, 2012). Each scenario in the data includes 3 hypothetical entry-level nursing jobs, which are differentiated by 12 characteristics and ranked from best to worst.

The initial analysis using MNL, ROL and HROL models detect a substantial degree of coefficient attenuation across ranks. For a further analysis, each model is specified as the kernel of a discrete mixture or latent class model. Latent class MNL and ROL tend to highlight the key features of unobserved heterogeneity in data well (Train, 2008; Keane and Wasi, 2012). Latent class HROL likewise can be expected to provide a good indicator of whether the conceptualised form of joint heterogeneity in the coefficients and rank-specific scale is present. The estimation results suggest that accounting for coefficient heterogeneity mitigates the extent of attenuation. This finding cannot be readily explained by supposing that respondents stated the best alternative with more certainty; then, the composite disturbance underlying the fixed coefficient models can be less heteroskedastic across ranks, due to the offsetting presence of the variance of omitted coefficient heterogeneity. Moreover, the latent class HROL estimates clearly suggest that the extent of attenuation is less when the coefficient configuration implies more deterministic behaviour. We thus describe how this form of joint heterogeneity can be conveniently imposed when adapting a very flexible mixed MNL model (Fiebig *et al.*, 2010) for rank-ordered data analysis.

The remainder of this paper is organised as follows. Section 2 summarises the problem of unstable coefficients and related models. Section 3 presents simulated evidence on the behaviour of ROL under pure stochastic misspecification, and discuss implications. Section 4 presents the analysis of the nursing job preferences data. Section 5 concludes.

## 2 The problem of unstable ROL coefficients

The usual cross-sectional setting for rank-ordered data is as follows. Agent $n \in \{1, 2, ..., N\}$ faces a choice set of $J_n > 2$ alternatives, where, for simplicity of notation, $J_n$ is assumed to equal $J$ for all $N$ agents as common in empirical applications.[2] For ease of presentation, the alternatives are assumed to be labelled arbitrarily and numerically so that equations can be written as if each agent faces the choice set $\mathbf{\Omega} = \{1, 2, \cdots, J\}$; this assumption is made without loss of generality because alternative $j \in \mathbf{\Omega}$ available to agent $n$ can be described by an agent-specific vector of $K$ observed characteristics, $\mathbf{x}_{nj}$. Each agent states which $H$ out of the $J$ alternatives she likes best where $1 < H \leq J-1$, and ordinally ranks these $H$ alternatives from best to worst without a tie.[3]

In the following formula, $r_{nh} \in \mathbf{\Omega}$ denotes agent $n$'s $h^{th}$ best alternative, and $\mathbf{\Omega}_{n,h-1}$ is a collection of $h-1$ alternatives she likes most. More specifically, $\mathbf{\Omega}_{n,h-1}$ refers to the empty set when $h = 1$ and $\cup_{i=1}^{h-1}\{r_{ni}\}$ when $2 \leq h \leq H$.

The rank-ordered logit (ROL) model specifies the probability of observing agent $n$'s ranking as:

$$P_n(\boldsymbol{\beta}) = \prod_{h=1}^{H} \frac{\exp(\boldsymbol{\beta} \cdot \mathbf{x}_{nr_{nh}})}{[\sum_{j \in \mathbf{\Omega} \setminus \mathbf{\Omega}_{n,h-1}} \exp(\boldsymbol{\beta} \cdot \mathbf{x}_{nj})]} \tag{1}$$

where $\boldsymbol{\beta}$ is a $K$-vector of parameters. The ROL formula is a product of multinomial logit (MNL) formulas; for model estimation, a single observation on agent $n$'s ranking is exploded into $H$ pseudo-observations on choices, according to Train's (p.157, 2009) parlance. The $h^{th}$ pseudo-observation is constructed as an independent observation on a choice among a set of alternatives excluding $\mathbf{\Omega}_{n,h-1}$. The sample size effectively increases $H$-fold, and $\boldsymbol{\beta}$ can be more precisely estimated than when each agent's best

[2]The following discussion can be easily adapted for cases where the number of alternatives varies across the agents, by making notations related to the choice set size agent-specific.

[3]In the special case when $H = J - 1$, all $J$ alternatives are effectively ranked from best to worst.

alternative among $\mathbf{\Omega}$ is observed alone. In the rest of this paper, the $h^{th}$ pseudo-choice data refer to the set of all agents' $h^{th}$ pseudo-observations.

From the perspective of microeconomic consumer theory, ROL describes the probability of (partially) observing a strict preference relation on $\mathbf{\Omega}$ which arises from the process of solving a random utility maximisation problem. Specifically, assume that agent $n$ obtains utility $U_{nj}$ from alternative $j \in \mathbf{\Omega}$:

$$U_{nj} = \boldsymbol{\beta} \cdot \mathbf{x}_{nj} + \varepsilon_{nj} \tag{2}$$

where the *ex-ante* random disturbance $\varepsilon_{nj}$ is iid type I extreme value (EV1) over alternatives.[4] At the time of decision making, $\varepsilon_{nj}$ is realised for each alternative, allowing all $J$ alternatives to be ranked unambiguously in descending order of realised utility indices. If agent $n$ chooses or states her first-ranked (ie utility-maximising) alternative, the probability of a choice response becomes MNL (McFadden, 1981), whereas if she states her top $H$-ranked alternatives, the probability of a rank-ordered response becomes ROL (Beggs *et al.*, 1981). (2) is called the logit utility function hereafter.

Each respondent's ranking can be always recoded as if $H$ had been smaller, and the ROL formula suggests that $\boldsymbol{\beta}$ may be consistently estimated by using any one of response variables detailing the top $Q$ ranks, where $1 \leq Q \leq H$. Its product structure implies that when the model is correctly specified, discarding some of available pseudo-choice datasets leads to only efficiency loss. Note that ROL reduces to MNL when the observed rankings are recoded as choices ($Q = 1$).

Several empirical studies since Chapman and Staelin (1982), however, have noticed what Foster and Mourato (2002) call 'the problem of unstable coefficients across ranks'. The ROL estimates tend to vary systematically as $Q$ varies. In particular, as Hausman and Ruud (1987) first emphasised, the estimated coefficients usually become attenuated monotonically as $Q$ is increased successively from 1 to $H$, as though worse-ranked alternatives have been stated more erratically.

Hausman and Ruud formulate the heteroskedastic rank-ordered logit (HROL) model that incorporates the key empirical regularity by modelling the probability correspond-

[4] Following McFadden (pp.205-206, 1981), our discussion associates the disturbance term with random fluctuations in the decision maker's state of mind concerning the utility she derives from each alternative. In stated preferences applications, all relevant attributes are observed, and it is less natural to describe the disturbance term as those attributes which are known to the decision maker but unobserved by the researcher.

ing to equation (1) as:

$$P_n(\boldsymbol{\beta}, \boldsymbol{\sigma}) = \prod_{h=1}^{H} \frac{\exp(\sigma_h \boldsymbol{\beta} \cdot \mathbf{x}_{nr_{nh}})}{[\sum_{j \in \boldsymbol{\Omega} \setminus \boldsymbol{\Omega}_{n,h-1}} \exp(\sigma_h \boldsymbol{\beta} \cdot \mathbf{x}_{nj})]} \tag{3}$$

where $\boldsymbol{\sigma} = [\sigma_2, \cdots, \sigma_H]$ is a vector of $H-1$ non-negative scalar parameters, with $\sigma_h$ measuring the scale of coefficients for the $h^{th}$ pseudo-choice data when $\sigma_1$ is normalised to 1. Maintaining that HROL is true, the pattern of attenuating coefficients can result from constraining all $\sigma_h$ to be identical, whereas in reality $1 > \sigma_2 > \cdots > \sigma_H$.

Hausman and Ruud also introduce a behavioural model from which HROL can be derived. This model parallels the top-down model of ranking behaviour from psychology (pp.69-70, Luce, 1959), and assumes that the observed ranking is a constructed response to the survey, instead of an *ex-ante* random preference relation on $\boldsymbol{\Omega}$ realised during the search for the utility-maximising alternative. Accordingly, it requires a more detailed statement about each agent's survey reporting behaviour.

Specifically, suppose that agent $n$ constructs her response by solving $H$ independent random utility maximisation problems in sequence. The choice set at the $h^{th}$ problem is $\boldsymbol{\Omega} \setminus \boldsymbol{\Omega}_{n,h-1}$, and the utility-maximising alternative in this set is ranked $h^{th}$ best in her response. The probability of a rank-ordered response becomes HROL if agent $n$ derives utility $U_{nj,h}$ from each alternative $j \in \boldsymbol{\Omega} \setminus \boldsymbol{\Omega}_{n,h-1}$:

$$U_{nj,h} = \boldsymbol{\beta} \cdot \mathbf{x}_{nj} + \varepsilon_{nj,h} / \sigma_h \tag{4}$$

where the disturbance $\varepsilon_{nj,h}$ is independent across $h$, and iid EV1 over alternatives for each $h$. Now the inequalities $1 > \sigma_2 > \cdots > \sigma_H$ can be said to hold when the agent states better-ranked alternatives with more certainty.

Over years, the problem of unstable coefficients has been mainly understood as a data problem, originating from manners in which individuals construct their responses sequentially. It has motivated many studies investigating the relative reliability of information on better- and worse-ranked alternatives (Foster and Mourato, 2002; Scarpa *et al.*, 2011) and that of rank-ordered and choice data (Boyle *et al.*, 2001; Caparros *et al.*, 2008). A driving concern appears to be that the reliability of rank-ordered responses may be compromised by the cognitive difficulty associated with making choices among inferior alternatives. Additional modelling approaches have also been proposed to describe the response construction sequence more generally. Ben-Akiva *et al.* (1992)

extend HROL by allowing a subset of coefficients to change disproportionately across the sequence, hypothesising that some attributes are not traded off to the same extent during earlier and later maximisation problems. Fok *et al.* (2012) model $\boldsymbol{\sigma}$ as a discrete random vector, speculating that different latent segments of respondents start making arbitrary decisions at different points in the sequence.

From the microeconomic perspective, however, the prevailing views on the origin of unstable coefficients appear rather unnatural. First, the problem of unstable coefficients in ROL implies that the logit utility function has been misspecified for an application at hand; it is not, by itself, a symptom of unreliable responses. Second, the microeconomic approach to rank-ordered data analysis postulates a random utility function, $U_{nj} = \boldsymbol{\beta} \cdot \mathbf{x}_{nj} + \varepsilon_{nj}$, and derives the probability of a particular preference relation consistent with random utility maximisation:

$$\Pr(U_{nr_{n1}} > U_{nr_{n2}} > \cdots > U_{nr_{nH}} > \max(U_{ni} \text{ for } i \in \boldsymbol{\Omega} \backslash \boldsymbol{\Omega}_{n,H}) | \mathbf{x}_{n1}, \mathbf{x}_{n2}, \cdots, \mathbf{x}_{nJ}) \quad (5)$$

where the disturbance terms $\boldsymbol{\varepsilon}_n = (\varepsilon_{n1}, \varepsilon_{n2}, \cdots, \varepsilon_{nJ})$ need not be iid extreme value, and other notations are as defined earlier. Invoking the inadequate modelling of a response construction sequence as a form of misspecification amounts to invalidating this approach generally, not only its special case involving the logit utility function that has produced the empirical problem of interest. The ROL model's product-of-MNL structure embodies the independence-of-irrelevant-alternatives (IIA) property induced by the iid extreme value disturbance terms. Most of econometric models consistent with (5) for some distributions of $\boldsymbol{\varepsilon}_n$ do not simplify to a product of multinomial choice models; they must be imposing wrong structures if a rank-ordered response needs to be modelled as a sequence of independently made choices in manner of (4).[5]

Section 3 explores the origin of unstable coefficients within the logit utility function, (2), itself without discrediting rank-ordered data and the microeconomic approach to analyse them. Two previous studies motivate the ensuing analysis. Hausman and Ruud (1987), in a less cited contribution, demonstrate how to estimate a subset of ROL coefficients using a procedure partially robust to stochastic misspecification in the

[5] For example, the rank-ordered probit model (p.158, Train, 2009) and the nested ROL model (Daksvik and Liu, 2009) can be derived by respectively assuming that $\boldsymbol{\varepsilon}_n$ follow a normal distribution and a correlated extreme value distribution. Now for simplicity, suppose that $J$=3. (5) can then be expressed as $Pr(U_{nr_{n2}} > U_{nr_{n3}} | \mathbf{x}_{n1}, \mathbf{x}_{n2}, \mathbf{x}_{n3}) \times Pr(U_{nr_{n1}} > \max(U_{nr_{n2}}, U_{nr_{n3}}) | \mathbf{x}_{n1}, \mathbf{x}_{n2}, \mathbf{x}_{n3}, U_{nr_{n2}} > U_{nr_{n3}})$. Conditioning on $U_{nr_{n2}} > U_{nr_{n3}}$ can be omitted when $\boldsymbol{\varepsilon}_n$ are iid extreme value (Beggs *et al.*, 1981) but not generally (p.157, Train, 2009).

logit utility function. Layton (2000) speculates that the problem may originate from the IIA property of the ROL model, though he does not investigate this issue in detail.

# 3 Implications of stochastic misspecification for ROL

## 3.1 Simulated evidence on misspecified disturbance to utility

The logit utility function, (2), can be misspecified in behaviourally important ways. The systematic component, $\boldsymbol{\beta} \cdot \mathbf{x}_{nj}$, ignores potential interpersonal variation in taste, due to unobserved heterogeneity in $\boldsymbol{\beta}$ and sometimes demographic interaction terms omitted from $\mathbf{x}_{nj}$. The unsystematic component, $\varepsilon_{nj}$, potentially ignores the residual parts of utility that are heteroskedastic and/or correlated over alternatives; such parts may be produced by misspecification in $\boldsymbol{\beta} \cdot \mathbf{x}_{nj}$ and/or exist naturally.

The logit utility function can be also subject to a subtle form of misspecification with minimal behavioural implications. In particular, the additive disturbance term may vary as an iid random variable, which is similar but not identical to an extreme value variable. At least this much of stochastic misspecification cannot be ruled out in an empirical study, when ROL is employed as a tractable approximation to an unknown data generating process (DGP).

Simulated examples in this subsection expose an empirically relevant facet of the ROL model which, to the best of our knowledge, has not been explicitly discussed in the literature. The problem of unstable coefficients can result from misspecifying the residual parts as iid extreme value, even when the true disturbance is almost iid extreme value and no other specification error is present. Each example uses 100 datasets generated from a specific DGP represented by a random utility function.

In Example 1, each dataset includes a random sample of 3000 agents who rank 5 different alternatives according to the following utility:

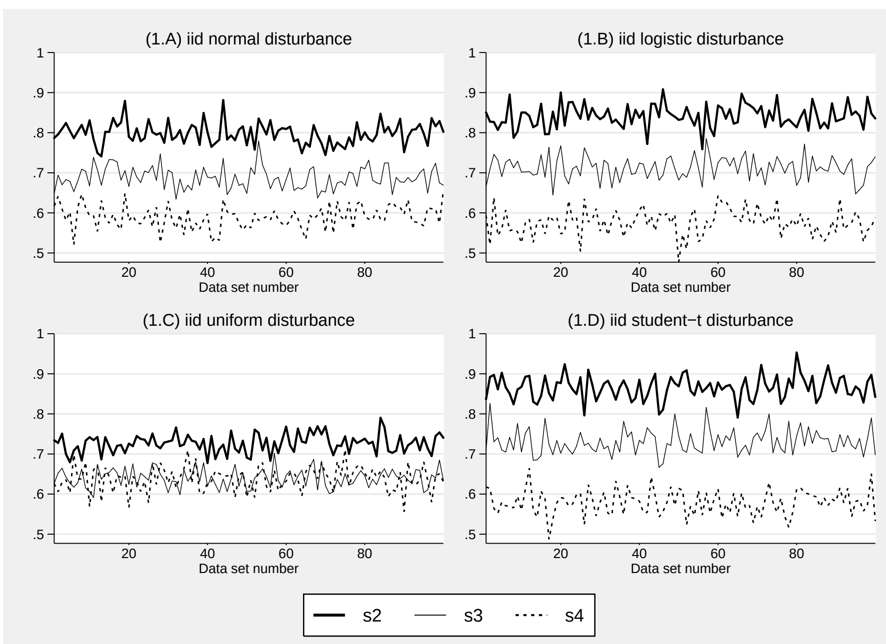
$$U_{nj} = x_{nj,1} + x_{nj,2} + e_{nj} \tag{6}$$
$$n = 1, 2, \cdots, 3000 \quad j = 1, 2, 3, 4, 5$$

where each of the observed attributes, $x_{nj1}$ and $x_{nj2}$, is generated from the standard normal distribution, and the unobserved disturbance $e_{nj}$ is iid normal with mean 0 and variance $\pi^2/6$. There is no correlation among the three terms. The observed response variable indicates each agent's ranking of all presented alternatives from best to worst (ie $H = 4$ as per notation in Section 2). ROL is a very slightly misspecified model for the resulting rank-ordered data, in the sense that the true disturbance distribution has a similar shape and the same variance as the EV1 distribution.

For each simulated sample, ROL is estimated using each of (recoded) response variables detailing the top $Q$ ranks where $Q$ varies from 1 to 4. The results over 100 datasets are summarised in Table 4 of Appendix 1; the estimated coefficients become successively attenuated across ranks, the average of each coefficient declining from 1.11 when $Q = 1$ to 0.89 when $Q = 4$.

Figure 1.A provides a related graphical summary. Here, HROL is estimated and the rank-specific scale parameters, $\boldsymbol{\sigma}$, are plotted separately for each dataset. Marker $\mathsf{s}h$ corresponds to $\sigma_h$. All estimated scales are significantly less than 1 (corresponding to no attenuation) at the 1% level, and suggest that the inequalities $1 > \sigma_2 > \sigma_3 > \sigma_4$ hold in every sample. The results look exactly as though simulated agents have stated better-ranked alternatives with more certainty, whereas in truth every one of them has stated an ordinal preference relation arising from random utility maximisation.

Figure 1: HROL scale estimates - Example 1 & its variants




Examples 2 and 3 are intended to show that the same problem occurs in a more empirically plausible environment. For either example, MNL is initially estimated using an actual dataset. Then, each artificial dataset is generated as a random sample of $N$

agents ranking $J$ alternatives according to the following utility:

$$U_{nj} = \mathbf{b} \cdot \mathbf{x}_{nj} + e_{nj} \qquad (7)$$
$$n = 1, 2, \cdots, N \quad j = 1, 2, \cdots, J$$

where the iid disturbance $e_{nj}$ is drawn from the same normal distribution as in Example 1, and the $K$-vectors $\mathbf{x}_{nj}$ and $\mathbf{b}$ respectively collect the observed attributes and the corresponding MNL coefficients. Within each new example, $N$, $J$, $K$ and $\mathbf{x}_{nj}$ are the same across all 100 artificial datasets. The observed response variable indicates each agent's ranking of all alternatives from best to worst ($H = J - 1$).[6]

Example 2 uses the dataset accompanying Stata module -mixlogit- (Hole, 2007), which is a subset of the electricity supplier choice data analysed in Huber and Train (2001). Here, $N = 1195$, $J = 4$ and $K = 6$; see Hole (2007) for a detailed data description. Example 3 uses the nursing job ranking data to be described and analysed in Section 4. The original responses are rank-ordered but recoded as choices for MNL estimation. Now, $N = 4208$, $J = 3$ and $K = 12$.

Figures 2.A and 3.A respectively plot the estimated HROL scales for Examples 2 and 3, all of which are again significantly less than 1. Moreover, Figure 2.A repeats that when $J \geq 3$, the attenuation pattern would look as though agents state better alternatives with more certainty; the inequalities $1 > \sigma_2 > \sigma_3$ hold in all but two ($45^{th}$ and $68^{th}$) of 100 replications. The ROL estimates from Examples 2 and 3 are respectively summarised in Table 5 and Table 6 of Appendix 1, and likewise show the pattern of attenuating coefficients across ranks.

When modelling multinomial choices, iid extreme value and iid normal disturbances often turn out to be empirically indistinguishable in terms of the resulting model behaviour. When modelling rank-ordered responses, their distinction can effect the well-known regularity suggestive of a specific response construction process because the maximum likelihood estimator (MLE) interprets pseudo-choice datasets as actual choice datasets. Put another way, when ROL or HROL is specified, MLE interprets a rank-ordered response as a sequence of independently made choices in manner of (4), with $\sigma_h$ being constrained to 1 for all $h$ in the case of ROL. But the rank-ordered probit (ROP) formula always involves a multivariate normal distribution function, which does not

[6] The actual datasets used here include repeated observations on the same respondents over different choice scenarios. Agent $n$ in equation (7) corresponds to a distinct pair of a respondent and a scenario, not a distinct respondent.

Figure 2: HROL scale estimates - Example 2 & its variants

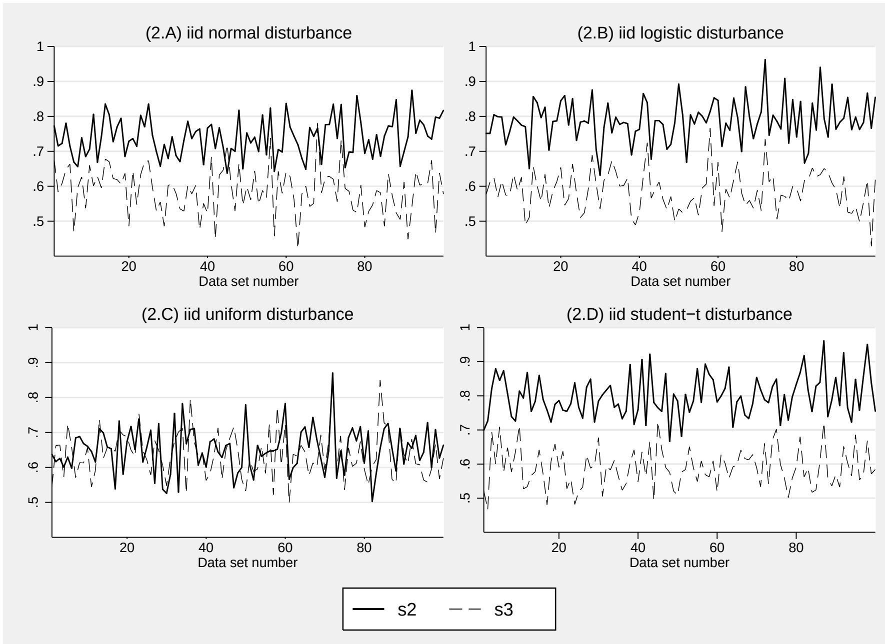
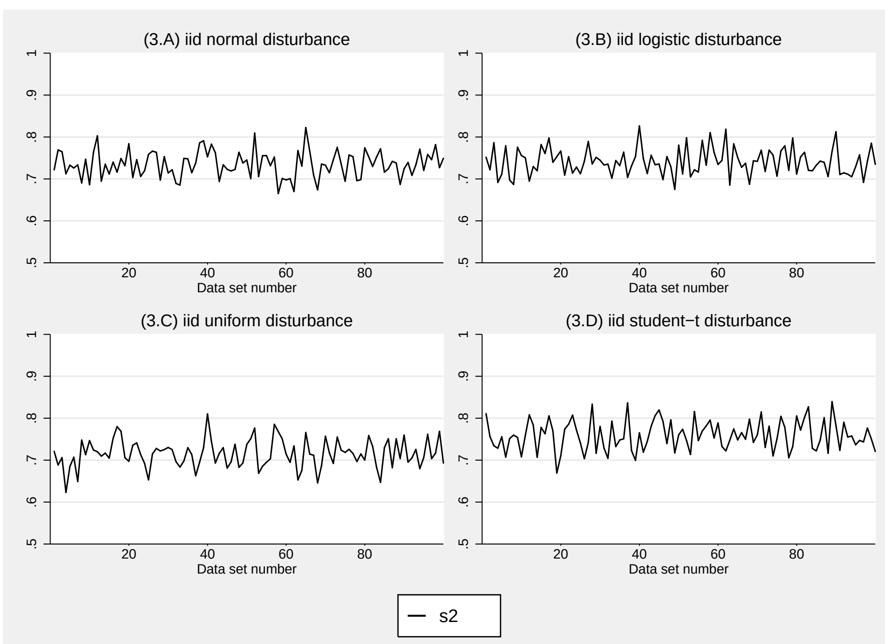



Figure 3: HROL scale estimates - Example 3 & its variants




simplify to a product of unconditional choice probabilities even when the disturbance is iid normal (p.158, Train, 2009). In consequence, the misspecified ROL model does not mimic the behaviour of the true ROP model directly; the ROL attempts to do so by mimicking that of a product of multinomial probit (MNP) models describing choices among successively smaller choice sets.

For a more specific discussion, consider Example 3 involving 3 alternatives per choice set. From MLE's perspective, each artificial dataset is a merged set of trinomial choice data (the first pseudo-choice data) and binomial choice data (the second pseudo-choice data), and there are trinomial and binomial probit models to be approximated by ROL. The trinomial probit coefficients are a suitably normalised version of **b** in DGP, and shared by a binomial probit model for randomly selected pairs of alternatives from the trinomial data. But the mimicked binomial probit model describes a choice between two least attractive alternatives; it is likely to have smaller coefficients because alternative pairs with a small difference in systematic utility components tend to be disproportionately included in the second pseudo-choice data, pushing the choice probabilities towards 0.5. Seeming heteroskedasticity across ranks results because the two mimicked MNP models do not share the same coefficients. This heuristic explanation can be extended to examples involving more alternatives.

ROL can be expected to exhibit unstable coefficients across ranks, not only when the true disturbance is iid normal. That a ranking probability can be specified as a product of sequential choice probabilities embodies the IIA property, that is not possessed by response probabilities derived from most disturbance distributions. It would be generally inappropriate to use one set of coefficients to describe how all multiplied choice probabilities vary as the observed attributes vary.

Figures 1-3 also plot the HROL scales estimated after replacing the normal disturbance distribution in each example with other commonly encountered distributions. The variance of each distribution has been set equal or similar to that of the EV1 distribution: logistic with the variance of $\pi^2/6$, uniform over $[-0.5^{0.5}\pi, 0.5^{0.5}\pi]$ and student-t with 5 degrees-of-freedom. Overall, these results again suggest that the problem of unstable coefficients needs not be tied to the unreliability of rank-ordered responses, or inadequacy of the microeconomic approach to rank-ordered data analysis. Since the uniform distribution has an equally probable and bounded support, it deviates more from the extreme value distribution than other bell-shaped distributions with unrestricted supports. Figures 1.C, 2.C and 3.C show that when the disturbance is iid uniform, the scale does not decline monotonically across ranks. The ROL estimates nevertheless become monotonically attenuated across ranks, because $\sigma_h$ is less than the unity for each $h \geq 2$.[7]

## 3.2 Interplay with preference heterogeneity

When analysing models derived from the logit utility function, (2), several researchers closely follow the empirical strategy described in Train (p.143, 2009). The systematic component, $\boldsymbol{\beta} \cdot \mathbf{x}_{nj}$, is specified to capture the parts of utility that are correlated or heteroskedastic over alternatives (preference heterogeneity hereafter),[8] through random parameterisation of $\boldsymbol{\beta}$ and augmentation of $\mathbf{x}_{nj}$ by suitable constant and interaction terms. The objective is to bring the variation of all residual parts reasonably close to that of an iid extreme value random variable.

This strategy allows describing the behaviour of utility-maximising agents more realistically, while partially exploiting the computational convenience of logit functional forms. The resulting mixed MNL model (McFadden and Train, 2000) is now well known, and the mixed ROL model has also been applied in both stated (Layton, 2000; Calfee *et al.*, 2001; Siikamaki and Layton, 2007; Train, 2008) and revealed (Train and Winston, 2007) preference analyses.

Subsection 3.1 highlights that the problem of unstable coefficients would be present unless the residual parts of utility are *exactly* iid extreme value. The associated results thus provide a useful background for discussing potential limitations and possible generalisations of a flexibly specified mixed ROL model.

It is now possible to provide analytic details to Layton's (2000) empirically motivated conjecture that mixed ROL would help mitigating the problem of unstable coefficients by relaxing the IIA property of ROL. To facilitate discussion, we recapitulate the sense in which the IIA property is relaxed. The *unconditional* ranking probability, $L_n(\boldsymbol{\theta})$, of a mixed ROL model generalising equation (1) is:

$$L_n(\boldsymbol{\theta}) = \int P_n(\boldsymbol{\beta}) f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta} = \int \left[ \prod_{h=1}^{H} \frac{\exp(\boldsymbol{\beta} \cdot \mathbf{x}_{nr_{nh}})}{[\sum_{j \in \boldsymbol{\Omega} \setminus \boldsymbol{\Omega}_{n,h-1}} \exp(\boldsymbol{\beta} \cdot \mathbf{x}_{nj})]} \right] f(\boldsymbol{\beta}|\boldsymbol{\theta}) d\boldsymbol{\beta} \tag{8}$$

[7] A summary of the ROL estimates for the non-normal disturbance examples is available upon request.

[8] Unobserved interpersonal heterogeneity in taste or $\boldsymbol{\beta}$ can induce such correlation and heteroskedasticity. The genuine covariances among the true disturbance terms can be accommodated as heterogeneity in $\boldsymbol{\beta}$ too, with a suitable augmentation of $\mathbf{x}_{nj}$ (McFadden and Train, 2000).

where density $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ specifies the population distribution of coefficients $\boldsymbol{\beta}$ using parameters $\boldsymbol{\theta}$. In this context, $P_n(\boldsymbol{\beta})$ is the ranking probability *conditional* on a particular $\boldsymbol{\beta}$. When $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ is degenerate with $\boldsymbol{\beta} = \boldsymbol{\theta}$, the original ROL is obtained; $L_n(\boldsymbol{\theta})$ and $P_n(\boldsymbol{\beta})$ are one and the same, both subject to IIA. When $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ is non-degenerate, $L_n(\boldsymbol{\theta})$ is no longer subject to IIA while $P_n(\boldsymbol{\beta})$ still is.

The last statement matters much more substantively than a similar statement concerning mixed MNL. The IIA property of $P_n(\boldsymbol{\beta})$ refers to the very product structure that render coefficients unstable across ranks when stochastic misspecification is present. In consequence, the use of mixed ROL does not obviate the problem, except when the residual parts were not iid extreme value initially only because of omitted preference heterogeneity, which is exactly described by the density $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ that the researcher has specified now.

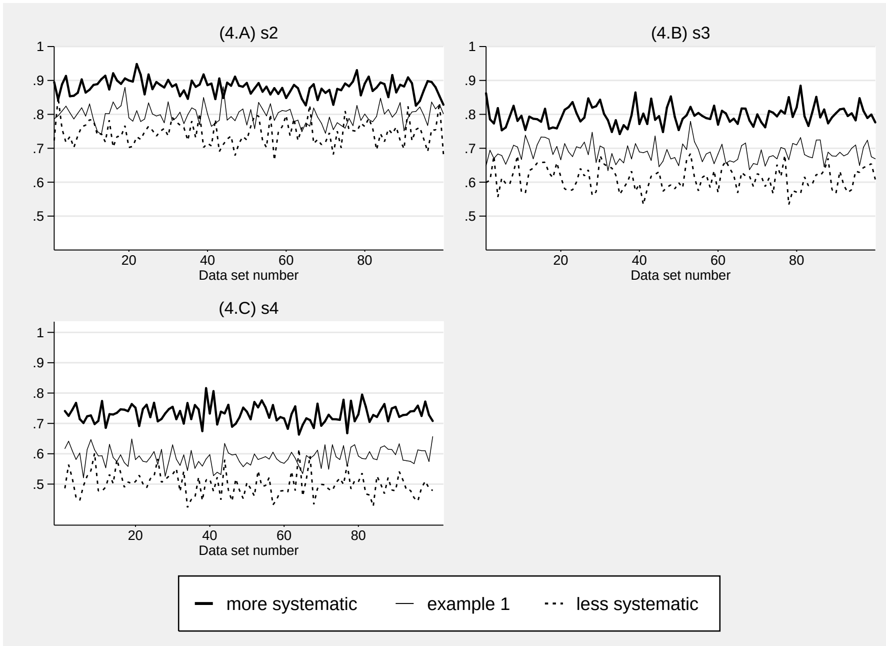
More often than not, mixed ROL can be expected to experience the problem of unstable coefficients. Not only there is little reason to expect the true disturbance to be iid extreme value, but also the exact modelling of preference heterogeneity is a practical impossibility due to computational and informational constraints. Any imperfectly captured heterogeneity, however minor, induces non-extreme value residuals, if only because a sum of random variables does not give an extreme value variable.

For a concrete example, suppose that the true random utility function is the same as in Example 1, except that the coefficients on $x_{nj1}$ and $x_{nj2}$ are draws from a multivariate normal distribution. A mixed ROL model with normal mixing would not mimic the behaviour of the true random coefficient ROP model directly; instead, it would mimic that of a sequence of random coefficient MNP models, using a single set of coefficient mean and covariance parameters, $\boldsymbol{\theta}$. The problem of unstable coefficients would now show up as attenuation of these parameters across ranks.

Nonetheless, mixed ROL can be generally expected to mitigate the problem in one important way. Incorporating preference heterogeneity increases the modelled parts of utility relative to the residual parts. Instability in coefficients across ranks, which results from misspecifying the latter as iid extreme value, would be lessened as the misspecified parts matter less for the postulated economic behaviour.

To illustrate the last point, Figure 4 plots the estimated HROL scales from Example 1, along with the corresponding estimates for two other variants of the same example. DGPs for these variants are identical to equation (6), except that the systematic component is larger ($2x_{nj,1} + 2x_{nj,2}$) and smaller ($0.5x_{nj,1} + 0.5x_{nj,2}$) respectively. For each

Figure 4: HROL scale estimates - Example 1 & its new variants




rank $h$ and across all 100 replications, $\sigma_h$ tends to be closer to 1 (no attenuation) when the simulated agent's random utility function becomes more deterministic.

Suppose now that the logit utility function has been flexibly specified so that the residual parts can be reasonably approximated as an iid extreme value variable. An important practical issue is whether coefficient attenuation due to potential approximation error may affect substantive conclusions from an analysis.

A mixed ROL derived from such utility function would suffice when the substantive results of interest do not depend on the scale of coefficients. For example, the researcher may be primarily interested in identifying the key patterns of preference heterogeneity (Train, 2008), or computing willingness-to-pay for different attributes (Calfee *et al.*, 2001). In all simulated datasets underlying Figures 1-4, the relative magnitudes of coefficients remain robust across different sets of ROL estimates obtained using different (recoded) response variables. Related details are available upon request but information in Appendix 1 is fairly suggestive. This form of robustness resonates with what Calfee *et al.* (2001) find in their rank-ordered data involving 13 alternatives and and our own empirical findings in subsection 4.2, on top of the well-known behaviour of binary logit estimates given stochastic misspecification (Cramer, 2007).

Coefficient attenuation across ranks is not innocuous when the analytic objective includes predicting the probability that an alternative is chosen or ranked first. From the microeconomic perspective, mixed ROL estimates can be plugged into the mixed MNL formula for demand forecasting, because both models are grounded in the same utility maximising behaviour. It is thus discomforting to know that slightly misspecifying the iid disturbance can bias choice probabilities towards $1/J$. Even when the forecasting objective is own and cross partial effects of attribute changes on choice probabilities, the present problem is far less innocuous than attenuation in binary logit coefficients due to neglected orthogonal regressors (Cramer, 2007). The usual MNL derivative expressions (p.58 & p.141, Train, 2009) suggest that when $J > 2$, decreased disparity among choice probabilities would not always offset the decreased scale of coefficients in such calculations.[9]

A mixed HROL model, which augments the preferred mixed ROL model with rank-scale specific parameters, $\boldsymbol{\sigma}$, may provide a convenient approach to address the biased forecasting issue. In all simulated datasets underlying Figures 1-4, the HROL coefficient estimates are practically identical to the MNL estimates (ie the ROL estimates using the first pseudo-choice data); again detailed results are omitted for brevity but Appendix 1 provides informative summary statistics.

Such mixed HROL would need to be specified with the view that $\boldsymbol{\sigma}$ captures the consequences of misspecifying the iid disturbance, instead of specific behavioural components. Thus, $\boldsymbol{\sigma}$ should be modelled as random parameters when coefficients in $\boldsymbol{\beta}$ are, because the joint distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ should have the property that for a realisation of $\boldsymbol{\beta}$ implying more deterministic choice behaviour, each $\sigma_h \in \boldsymbol{\sigma}$ is closer to 1; the misspecified residual parts matter less for agents with such preferences. Section 4 introduces how this modelling implication can be implemented within discrete (subsection 4.1) and continuous (subsection 4.2) mixture contexts.

The mixed HROL model described thus far is fundamentally different from the HROL model with non-random $\boldsymbol{\beta}$ and random $\boldsymbol{\sigma}$ due to Fok *et al.* (2012). In their model, $\boldsymbol{\sigma}$ is a discrete random vector. Each of its mass points has a distinct number of leading 1s followed by 0s, and represents a latent segment following a particular response construction process (see Section 2). When the microeconomic perspective is maintained as in our analysis, there is no reason to impose such selection of mass points (see Figures 1-4), and consider heterogeneity of $\boldsymbol{\sigma}$ in isolation from that of $\boldsymbol{\beta}$.

[9]For example, given a realised $\boldsymbol{\beta}$, coefficient attenuation leads to attenuation in own partial effects when the choice probability would lie between $1/J$ and $1/2$ without coefficient attenuation.

The mixed HROL model has a conceptual limitation. While it accounts for the implication of an almost inevitable form of misspecification in the logit utility function, it cannot be directly derived from any utility function using the microeconomic approach. Another possibility which does not have this limitation is to estimate mixed MNL using the first pseudo-choice data. But when estimating random coefficient models, such approach may not only entail efficiency loss but also a reduction in model flexibility. The number of repeated observations per individual crucially affects both theoretical and empirical identification of mixed logit model parameters (Walker *et al.*, 2007; Hess and Train, 2011), and an MNL version of the preferred mixed ROL model may not be estimable; rank-ordered data effectively provide $H$ times as many repeated observations as recoded choice data. Subsection 4.1 reports a related empirical finding.

# 4 Empirical analysis

Section 3 suggests that the problem of unstable coefficients can be conceptualised as resulting from stochastic misspecification in the logit utility function, (2). Subsection 3.2 has discussed the related modelling implications and expected behaviour of estimates.

This section investigates how such expectation plays out empirically, using rank-ordered data collected as part of an ongoing longitudinal survey of Australian nursing students and graduates. Kenny *et al.* (2012) review the first wave of the survey in detail. Doiron *et al.* (2011) and Yoo and Doiron (2012) use these data with distinct objectives from the present analysis. In line with the main objective of the survey, the first paper focuses on discovering the key determinants of nursing job choices to formulate informed policy recommendations. The latter study examines the comparability of preferences elicited by a best-worst alternative experiment and a best-worst attribute-level experiment.[10]

## 4.1 Data and methods

The estimation sample consists of 526 respondents who completed an online survey between September 2009 and September 2010. They were recruited from the Bachelor of Nursing (BN) degree students enrolled at two large Australian universities during 2008-

[10]The best-worst alternative experiment is described shortly. In the best-worst attribute-level experiment, respondents see one nursing job described by several characteristics set at specific levels, and state the best characteristic and the worst characteristic.

2010: the University of Technology Sydney and the University of New England. The sample consists of nursing students in each year of the 3-year program and graduates within 12 months of completion.

In the stated preference component of the survey, each respondent faces 8 different scenarios or choice sets. Each scenario consists of 3 hypothetical jobs -labelled A, B and C- differentiated by salary and eleven non-salary attributes. The respondent then states the best job and the worst job of each presented scenario, effectively ranking all 3 jobs from most to least preferred.[11]

Table 1 lists 4 different levels of salary and 2 different levels of each non-salary attribute used for the job specification. The selection of attributes reflect characteristics that have been shown to matter in the quitting decision and job satisfaction of nurses (Seago *et al.*, 2001; Naude and McCabe, 2005). The levels of the attributes reflect those found in entry-level jobs as registered nurses in Australia. The feedback from an earlier pilot study involving 60 students indicates that the attributes and levels are appropriate for the intended context.

The scenarios are constructed from an initial set of 16 jobs which form a resolution 3 fractional factorial design. The other two jobs in each scenario are determined by the addition of two generators, chosen so that the resulting set of 16 scenarios of size 3 is D-optimal when all coefficients in the standard MNL model are zero. Two sets of 16 scenarios are constructed using two different resolution 3 fractions so that a larger proportion of the sample space is covered. Each set is divided into two subsets or versions of 8 scenarios, and each person is randomised to one of the resulting 4 versions.

We now redefine notations to suit these data. Respondent $n \in \{1, 2, \cdots, N\}$ ranks 3 alternatives in $\mathbf{\Omega} = \{A, B, C\}$ over $T$ different scenarios. Alternative $j \in \mathbf{\Omega}$ in scenario $t \in \{1, 2, \cdots T\}$ presented to respondent $n$ is described by vector $\mathbf{x}_{njt}$ collecting $K$ attributes. $r_{nth} \in \mathbf{\Omega}$ denotes respondent $n$'s $h^{th}$ best alternative in scenario $t$. In the present context, $N = 526$, $T = 8$ and $K = 12$.

$\mathbf{x}_{njt}$ includes a binary indicator of one level of each non-salary attribute (11 in total) and the natural logarithm of salary.[12] As the job labelling is arbitrary, most, if not all, of covariances in utility over alternatives would operate through heterogeneous tastes for these attributes.

[11]This elicitation format has been advocated as being cognitively easier than directly asking respondents to rank alternatives; see Scarpa *et al.*(2011).

[12]The logarithmic transformation allows capturing parsimoniously the decreasing marginal utility of salary, found in a preliminary analysis involving three binary salary-level indicators.

Table 1: Job attributes and associated levels

| **Glossary definition of attribute** | **Attribute name** | **Levels** | **Variable** |
|---|---|---|---|
| The type of hospital where the new graduate program is located | Location | Private hospital<br>Public hospital | (Omitted)<br>public hosp |
| The number of rotations to different clinical areas | Clinical rotations | None<br>Three | (Omitted)<br>3 rotations |
| Whether the new graduate program offers fulltime and part-time positions, or fulltime only | Work hours | Fulltime only<br>Part-time or fulltime | (Omitted)<br>flex hours |
| The flexibility of the rostering system in accommodating requests | Rostering | Inflexible, does not allow requests<br>Flexible, usually accommodating requests | (Omitted)<br>flex rost |
| The hospital's reputation regarding staffing levels | Staffing levels | Frequently short of staff<br>Usually well-staffed | (Omitted)<br>well staff |
| The hospital's reputation regarding the workplace culture in terms of support from management and staff | Workplace culture | Unsupportive management and staff<br>Supportive management and staff | (Omitted)<br>supp mgt |
| The hospital's reputation regarding the physical work environment in terms of equipment and appearance | Physical environment | Poorly equipped and maintained facility<br>Well equipped and maintained facility | (Omitted)<br>well equip |
| The hospital's reputation regarding whether nurses are encouraged and supported in professional development and career progression | Professional development and progression | No encouragement for nurses<br>Nurses encouraged | (Omitted)<br>encourage |
| The parking facilities | Parking | Limited<br>Abundant and safe | (Omitted)<br>abund park |
| The hospital's reputation regarding the responsibility given to nurses, relative to their qualifications and experience | Responsibility | Too much responsibility<br>Appropriate responsibility | (Omitted)<br>app resp |
| The hospital's reputation regarding the quality of patient care | Quality of care | Poor<br>Excellent | (Omitted)<br>excell care |
| The gross weekly salary | Salary | $800<br>$950<br>$1,100<br>$1,250 | log salary |

Each rank-ordered response $\{r_{nt1}, r_{nt2}\}$ is assumed to reflect a preference relation on $\mathbf{\Omega}$ realised from maximising random utility $U_{njt} = \boldsymbol{\beta} \cdot \mathbf{x}_{njt} + \varepsilon_{njt}$, where $\boldsymbol{\beta}$ is a $K$-vector of utility weights. The random disturbance $\varepsilon_{njt}$ is assumed to be iid EV1, as an approximation to an unknown disturbance distribution. To account for potential approximation error, the probability of observing respondent $n$'s responses over $T$ scenarios is specified as HROL:

$$P_n(\boldsymbol{\beta}, \sigma) = \prod_{t=1}^{T} \frac{\exp(\boldsymbol{\beta} \cdot \mathbf{x}_{nr_{nt1}t})}{[\sum_{j \in \mathbf{\Omega}} \exp(\boldsymbol{\beta} \cdot \mathbf{x}_{njt})]} \frac{\exp(\sigma\boldsymbol{\beta} \cdot \mathbf{x}_{nr_{nt2}t})}{[\sum_{j \in \mathbf{\Omega} \setminus \{r_{nt1}\}} \exp(\sigma\boldsymbol{\beta} \cdot \mathbf{x}_{njt})]} \tag{9}$$

for a given $\boldsymbol{\beta}$ and scalar parameter $\sigma$, where the latter accommodates a scale change in the second pseudo-observations. Such change can result from exploding the data despite stochastic misspecification.

To capture interpersonal taste variation, $\boldsymbol{\beta}$ are treated as random parameters. Since the misspecified disturbance matters to varying extents for individuals with different preferences, $\sigma$ is also specified as a random parameter. Density $g(\boldsymbol{\beta},\sigma|\boldsymbol{\delta})$ describes the joint distribution of $\boldsymbol{\beta}$ and $\sigma$ as a function of $\boldsymbol{\delta}$. The current literature tends to investigate heterogeneity in $\boldsymbol{\beta}$ (eg. Calfee *et al.*, 2001 ) and in $\sigma$ (Fok *et al.*, 2012) separately, depending on whether the microeconomic approach or the constructed response approach is taken.

The unconditional probability of observing respondent $n$'s responses, $L_n(\boldsymbol{\delta})$, is obtained by integrating $P_n(\boldsymbol{\beta}, \sigma)$ over $g(\boldsymbol{\beta},\sigma|\boldsymbol{\delta})$:

$$L_n(\boldsymbol{\delta}) = \int \int P_n(\boldsymbol{\beta}, \sigma) g(\boldsymbol{\beta}, \sigma|\boldsymbol{\delta}) d\boldsymbol{\beta} d\sigma \tag{10}$$

and maximised with respect to $\boldsymbol{\delta}$.

To obtain a tractable functional form, $g(\boldsymbol{\beta},\sigma|\boldsymbol{\delta})$ is specified as a discrete distribution with $C$ mass points at ($\boldsymbol{\beta}_c$, $\sigma_c$) for $c = 1, 2, \cdots, C$. The relative frequency of each point is denoted $\pi_c$, where $0 < \pi_c < 1$ and $\sum_{c=1}^{C} \pi_c = 1$. Adopting the parlance for a latent class model, each point corresponds to a class and $\pi_c$ is the population share of class $c$. Train (2008) highlights that discrete mixture models have a non-parametric approximation property in relation to an arbitrary mixture model. Available evidence shows that discrete mixture or latent class ROL (Train, 2008) and MNL (Keane and Wasi, 2012) models tend to provide a good summary of unobserved heterogeneity in data.

The unconditional probability becomes:

$$L_n(\boldsymbol{\delta}) = \sum_{c=1}^{C} \pi_c P_n(\boldsymbol{\beta}_c, \sigma_c) \tag{11}$$

where the parameters to be estimated, $\boldsymbol{\delta}$, include $\boldsymbol{\beta}_c$, $\sigma_c$ and $\pi_c$ for each class, except $\pi_1$ which is set to $1-\sum_{c=2}^{C} \pi_c$ for identification. (11) is called latent class HROL (LHROL) hereafter.

LHROL is both computationally simple and analytically useful. Even with a large $C$, it can be easily estimated by adapting Train's (2008) EM algorithm for latent class MNL, by replacing the MNL kernel with the HROL kernel. Since each class has its own $\boldsymbol{\beta}_c$ and $\sigma_c$, it is also easy to inspect whether $\sigma$ is indeed closer to 1 when $\boldsymbol{\beta}$ implies more systematic behaviour. Note that LHROL does not impose, but allows data to speak if the conceptualised form of joint heterogeneity in $\boldsymbol{\beta}$ and $\sigma$ is present.

Other models analysed below can be viewed as special cases of (11). C-LHROL is obtained by constraining $\sigma_c$ to be identical across classes. Latent class ROL (LROL) is obtained by further constraining $\sigma_c$ to 1 for each $c$. LHROL gives the same log-likelihood as the usual latent class MNL (LMNL) up to a constant summand when $\sigma_c = 0$ for each $c$.[13] MNL, ROL and HROL are nested in their namesake latent class models, occurring when only one class is specified ($C = 1$).

To choose the number of classes, we follow a procedure similar to Train's (2008). Each model is estimated several times via the EM algorithm, with a successively increasing number of classes from 2 through 10, and the number that yields the smallest Bayesian Information Criterion (BIC) is selected for the final specification. The following analysis specifies 3 classes for LMNL and 4 classes for other models (LROL, C-LHROL and LHROL).

The final results have been obtained via the Newton algorithm, using the EM estimates as starting values.[14] Each model's Hessian becomes singular when classes increase beyond the BIC optimal number found in the EM procedure, even when alternative gradient-based algorithms and convergence criteria are used. The immediate implication is that moving from mixed ROL (LROL) to mixed MNL (LMNL) involves a reduction in model flexibility in the present application.[15]

[13]We do not include this summand, $4208 \times \log(0.5)$, in the reported log-likelihood and BIC.

[14]All latent class models have been estimated using TSP International 5.1. Basic MNL, ROL and HROL have been estimated using Stata 11.2/IC.

[15]Also see the end of subsection 3.2.

## 4.2 Main findings[16]

To anticipate results from models accounting for preference heterogeneity, we initially analyse MNL, ROL and HROL. The first three columns of Table 2 report corresponding estimates. A severe problem of unstable coefficients is present, despite the survey presents only 3 jobs per choice set, and adopts the best-worst elicitation format to reduce cognitive burden on respondents.

All MNL and ROL coefficients are precisely estimated yet mostly disagree on the first significance figures, exhibiting the usual attenuation pattern across ranks. The HROL scale, $\sigma$, likewise suggests that the coefficient vector, $\boldsymbol{\beta}$, is scaled by 0.563 in the second rank. A behavioural interpretation in accordance with (4) would suggest that the disturbance variance increases 3 times ($0.563^{-2} > 3$) across the response construction sequence.

From the perspective of Section 3, however, there is little reason to find the MNL estimates more credible than the ROL estimates. Both MNL and ROL are consistent with the same logit utility function with homogeneous coefficients, and coefficient attenuation can result even when this utility function is only slightly misspecified.

The last three columns of Table 2 report the proportion of salary a person is willing to give up to obtain each attribute-level. It is computed as $[\exp(b_k/b_{\log salary}) - 1]$ where $b_k$ is the point estimate of the coefficient on attribute $k$, and hence should be the same across ranks when $b_k/b_{\log salary}$ behaves as in the datasets underlying Figure 3. Some variation exists across the three models, suggesting that it is not only the iid disturbance that has been misspecified. Nonetheless, the extent of variation never exceeds 3 percentage points, except excell care (excellent quality of care) for which the MNL and ROL results are 0.34 and 0.39 respectively.

Figure 5 plots the probabilities of choosing the first-ranked alternatives (choice probabilities hereafter) predicted by ROL and HROL against those predicted by MNL. The former two predictions would lie on the 45 degree line if they coincide with the MNL predictions. Due to coefficient attenuation, the ROL prediction tends to lie above (below) this line when the MNL prediction is less (greater) than 1/3. The HROL estimates are adjusted for attenuation and resulting predictions are clustered around the 45 degree line, almost coinciding with the MNL predictions in all 4208 choice sets.

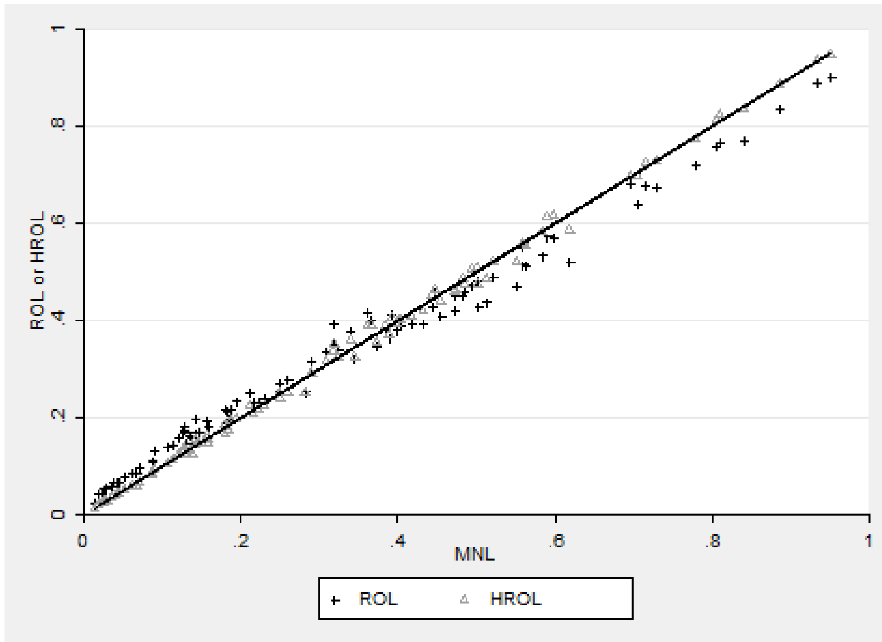
[16] All statistical tests in this subsection have been conducted at the 1% level. As we use designed survey data, the model estimates tend to be very precise.

Table 2: Detailed estimation results - basic models

| | Parameter estimates | | | Willingness-to-pay | | |
|---|---|---|---|---|---|---|
| | MNL | ROL | HROL | MNL | ROL | HROL |
| public hosp | 0.239 | 0.130 | 0.204 | 0.085 | 0.059 | 0.075 |
| | (0.039) | (0.028) | (0.034) | | | |
| 3 rotations | 0.212 | 0.153 | 0.182 | 0.076 | 0.070 | 0.067 |
| | (0.039) | (0.027) | (0.034) | | | |
| flex hours | 0.138 | 0.0756 | 0.119 | 0.049 | 0.034 | 0.043 |
| | (0.036) | (0.027) | (0.032) | | | |
| flex rost | 0.568 | 0.390 | 0.526 | 0.202 | 0.177 | 0.192 |
| | (0.043) | (0.031) | (0.039) | | | |
| well staff | 0.409 | 0.360 | 0.426 | 0.146 | 0.164 | 0.156 |
| | (0.037) | (0.029) | (0.034) | | | |
| supp mgt | 1.032 | 0.828 | 1.024 | 0.368 | 0.377 | 0.375 |
| | (0.049) | (0.039) | (0.049) | | | |
| well equip | 0.375 | 0.349 | 0.391 | 0.134 | 0.159 | 0.143 |
| | (0.039) | (0.029) | (0.035) | | | |
| encourage | 0.545 | 0.440 | 0.558 | 0.194 | 0.200 | 0.204 |
| | (0.045) | (0.034) | (0.042) | | | |
| abund park | 0.0809 | 0.0738 | 0.0963 | 0.029 | 0.034 | 0.035 |
| | (0.038) | (0.028) | (0.033) | | | |
| app resp | 0.462 | 0.376 | 0.460 | 0.165 | 0.171 | 0.169 |
| | (0.046) | (0.035) | (0.043) | | | |
| excell care | 0.822 | 0.730 | 0.858 | 0.293 | 0.332 | 0.314 |
| | (0.049) | (0.038) | (0.046) | | | |
| log salary | 2.807 | 2.199 | 2.732 | | | |
| | (0.170) | (0.126) | (0.164) | | | |
| $\sigma$ | | | 0.563 | | | |
| | | | (0.031) | | | |
| log-likelihood | -3497.6 | -6211.9 | -6149.0 | | | |
| bic | 7108.5 | 12543.3 | 12427.4 | | | |
| parameters | 12 | 12 | 13 | | | |

Standard errors in parentheses have been adjusted for clustering at the respondent level. All utility weights are significant at the 1% level. $\sigma$ is significantly less than the unity at the 1% level.

Figure 5: Choice probabilities predicted by basic models




Overall the basic models show a fair amount of agreement on substantive results, despite omitted preference heterogeneity. We take this as a sign that there is no other behaviourally important misspecification in the utility function, and proceed to an analysis of latent class models (LMNL, LROL, C-LHROL, LHROL).[17] These models involve several class-specific parameters, without direct correspondence among classes across models. For succinct comparisons, we focus on the weighted average estimates using class shares as the weights, graphical comparisons of identified preference segments, and the in-sample predictive performance of each model.

Table 3 reports the average estimates from each latent class model. LROL obviates the IIA property of the unconditional probability but maintains that of the conditional probability. More often that not, it would experience coefficient attenuation (see subsection 3.2), and as expected, the average LROL estimates tend to have the smallest magnitudes. More directly, the estimated C-LHROL scale, $\sigma$, is 0.615 and significantly less than 1. The average scale in LHROL is similar.

Accounting for preference heterogeneity, however, mitigates coefficient attenuation found in ROL via two different routes in the present application. First, the misspecified residual parts matter relatively less now, making $\sigma$ increase from 0.563 in HROL to 0.615

[17]Had the substantive results disagreed a lot, it would have been worthwhile investigating more sophisticated parametric form specifications of the systematic utility component first.

Table 3: Average parameter estimates - latent class models

| | LMNL | LROL | C-LHROL | LHROL |
|---|---|---|---|---|
| public hosp | 0.288 | 0.164 | 0.250 | 0.242 |
| | (0.049) | (0.050) | (0.052) | (0.051) |
| 3 rotations | 0.307 | 0.235 | 0.318 | 0.291 |
| | (0.050) | (0.052) | (0.054) | (0.055) |
| flex hours | 0.195 | 0.104 | 0.151 | 0.150 |
| | (0.045) | (0.032) | (0.039) | (0.038) |
| flex rost | 0.660 | 0.453 | 0.579 | 0.578 |
| | (0.048) | (0.034) | (0.043) | (0.042) |
| well staff | 0.499 | 0.413 | 0.508 | 0.486 |
| | (0.048) | (0.035) | (0.042) | (0.041) |
| supp mgt | 1.129 | 1.116 | 1.393 | 1.323 |
| | (0.054) | (0.071) | (0.094) | (0.093) |
| well equip | 0.497 | 0.461 | 0.578 | 0.550 |
| | (0.052) | (0.053) | (0.058) | (0.059) |
| encourage | 0.587 | 0.550 | 0.679 | 0.659 |
| | (0.046) | (0.040) | (0.056) | (0.052) |
| abund park | 0.117 | 0.120 | 0.150 | 0.149 |
| | (0.046) | (0.033) | (0.040) | (0.039) |
| app resp | 0.553 | 0.538 | 0.626 | 0.618 |
| | (0.050) | (0.047) | (0.054) | (0.050) |
| excell care | 1.003 | 0.926 | 1.113 | 1.089 |
| | (0.066) | (0.050) | (0.067) | (0.066) |
| log salary | 3.153 | 2.695 | 3.280 | 3.102 |
| | (0.223) | (0.173) | (0.220) | (0.204) |
| $\sigma$ | | | 0.615 | 0.619 |
| | | | (0.032) | (0.035) |
| log-likelihood | -3315.88 | -5785.35 | -5738.36 | -5730.03 |
| bic | 6869.84 | 11890.24 | 11802.5 | 11804.64 |
| parameters | 38 | 51 | 52 | 55 |

Standard errors in parentheses have been adjusted for clustering at the respondent level. All average utility weights are significant at the 1% level $\sigma$ is significantly less than the unity at the 1% level.

in C-LHROL. This increase is equivalent to a 16% decrease in the heteroskedasticity of the disturbance variance across ranks, if we adopt the behavioural interpretation of $\sigma$ momentarily to aid interpretation. Second, LROL is subject to a smaller error in approximating the unknown taste distribution than LMNL for which one fewer class is empirically identified. The relative disparity of the average LROL and LMNL estimates is thus less pronounced than that of the basic ROL and MNL estimates.

That $\sigma$ is larger in C-LHROL than in HROL deserves a further comment. This result is just as expected from subsection 3.2. It cannot be, however, readily explained by conceptualising instead that respondents construct responses in accordance with (4). The residual variance for HROL would then comprise the variance of the true additive disturbance that is heteroskedastic across ranks, and that of omitted preference heterogeneity. The latter mitigates heteroskedasticity as long as it is not larger at the second maximisation problem, and is likely to be smaller for the second problem involving one fewer nursing job. As a result, accounting for preference heterogeneity can be expected to lead to a decline, instead of an increase, in $\sigma$.

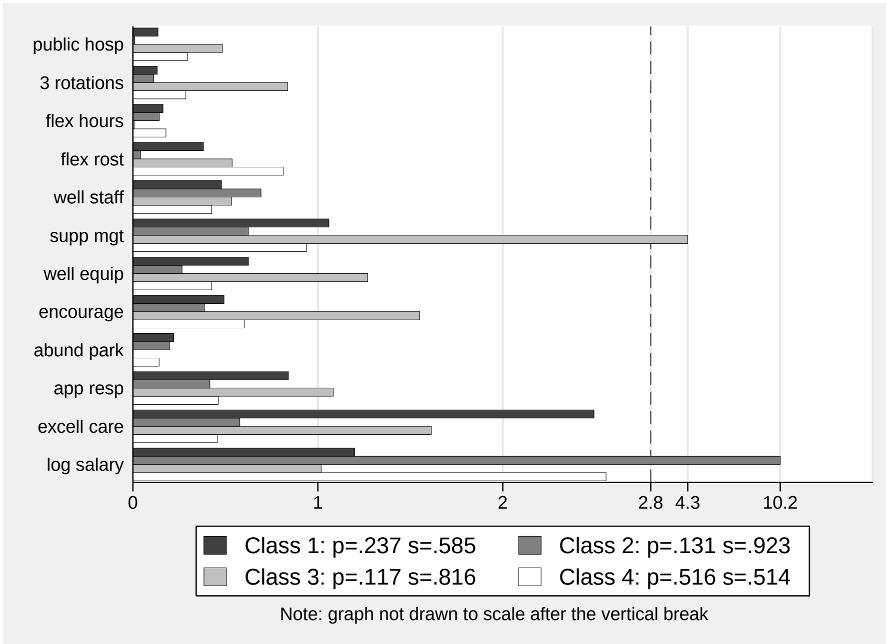
Figure 6 plots class-specific utility weights from LHROL. Statistics p and s next to each class label respectively report the population share ($\pi_c$) and the rank-specific scale ($\sigma_c$) of that class. Table 7 of Appendix 2 provides detailed estimation results for all latent class models.

All estimated scales are statistically and practically larger than 0, the smallest being 0.514 (Class 4). In other words, no class exhibits the lack of ranking capabilities conceptualised by Fok *et al.* (2012). Instead, the results empirically illustrate the form of joint heterogeneity conceptualised in subsection 3.2; $\sigma_c$ tends to be closer 1 when $\boldsymbol{\beta}_c$ implies more deterministic preferences.

The degree of coefficient attenuation is the least for Classes 2 and 3, the scales of which are 0.923 and 0.816 respectively. Respondents in Class 2 are very salary-sensitive, with an extremely large coefficient on log salary. Their responses can be very easily predicted using the observed attributes, as they would tend to rank jobs in order of salary levels. Respondents in Class 3 rank jobs on the basis of trade-offs among different attributes, to a larger extent than respondents in Classes 1 and 4. Several utility weights, on supp mgt (supportive management) in particular, are much larger for Class 3, while others are similar or not as smaller.

Class 4 experiences the largest attenuation ($\sigma_4 = 0.514$) followed by Class 1 ($\sigma_1 = 0.585$). The residual parts matter the most for Class 4. While this class has the second largest utility weight on log salary (2.558), it is not large enough to make up for the

Figure 6: LHROL utility weight estimates




other small weights; the utility index would change only by 1.142 in response to the maximum possible salary increase from \$800 to \$1250. Unless the choice set contains a clearly superior job and a clearly inferior job, it is relatively hard to predict which jobs respondents in Class 4 would rank best and worst. The utility weights for Class 1 also tend to be small, except the weight on excell care (2.493) which is the largest among all classes. Their behaviour is somewhat easier to predict because, given a set of jobs similarly attractive otherwise, they are highly likely to rank jobs with excell care = 1 better when available.

Figures 7 and 8 respectively plot class-specific utility weights from LROL and LMNL. The C-LHROL results are omitted because they deviate a little from the LHROL estimates, even though the underlying constraints ($\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$) are rejected using a LR test statistic (16.66).

The estimated pattern of preference heterogeneity is remarkably robust across LHROL and LROL. Each coefficient vector, $\boldsymbol{\beta}_c$, in LROL looks almost like an attenuated version of a coefficient vector in LHROL. Put another way, it would not be overly wrong to summarise that LHROL scales up the given LROL estimates so that the new estimates are more suitable for forecasting purposes.

Figure 7: LROL utility weight estimates

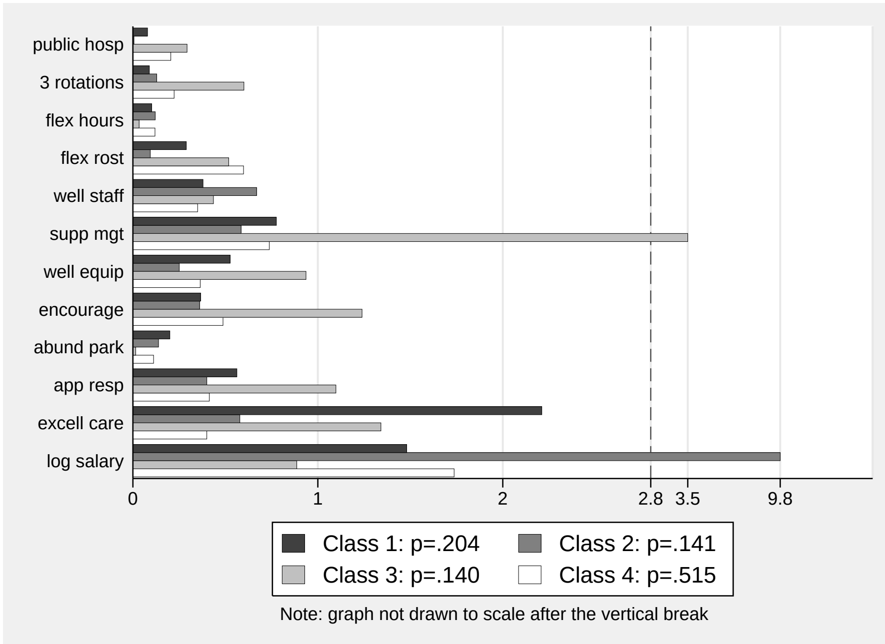
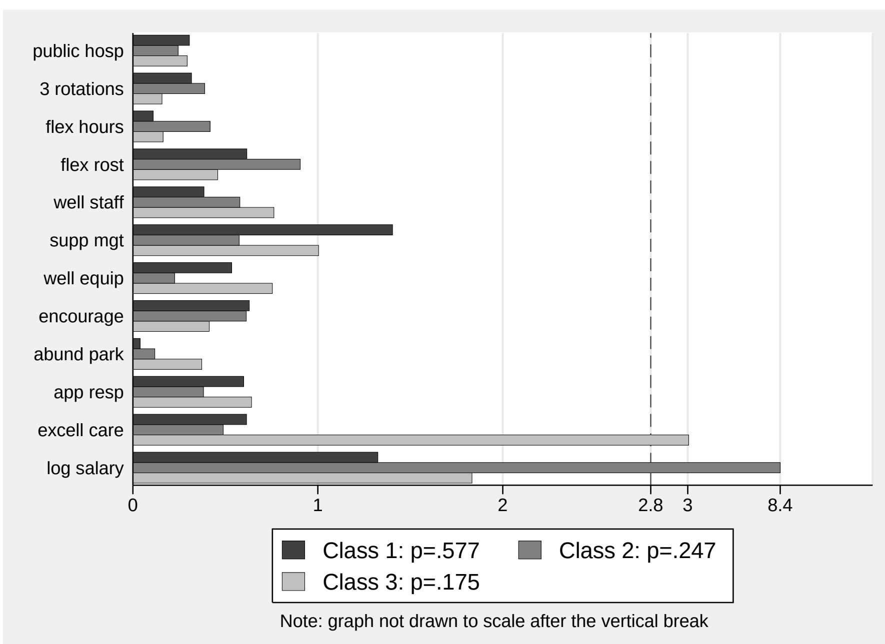
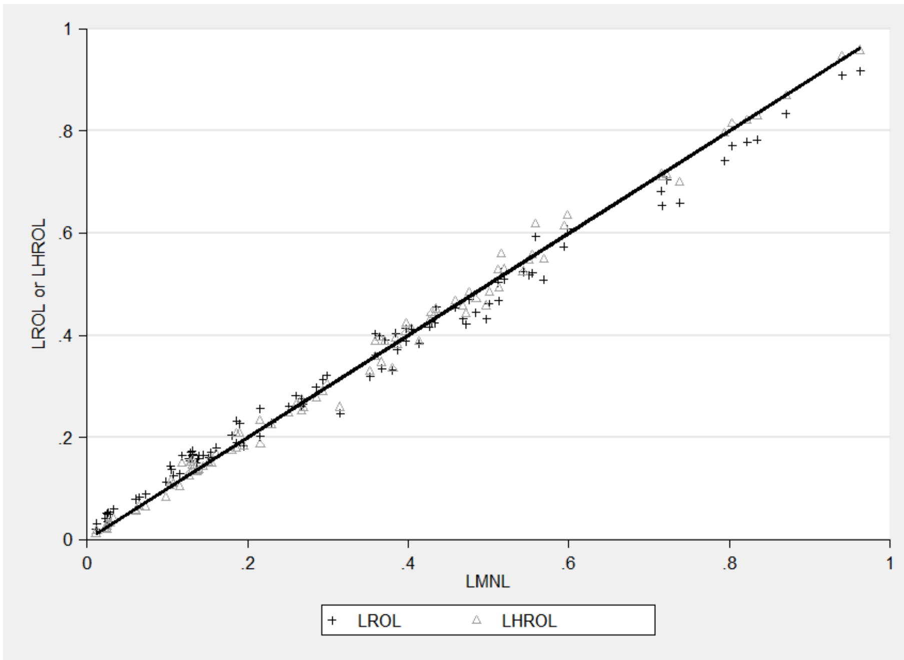



Figure 8: LMNL utility weight estimates




LMNL relatively understates the amount of preference heterogeneity as one fewer class is empirically identified without using the second pseudo-choice data. A major

Figure 9: Choice probabilities predicted by latent class models




consequence seems to be the loss of clear distinction between respondents who trade off among several attributes more and less systematically, like Classes 3 and 4 in LHROL.

Figure 9 is a latent class model analogue to Figure 5. It plots unconditional (that is, class share weighted average) choice probabilities predicted by LROL and LHROL against those predicted by LMNL. As the LMNL predictions deviate from 1/3, the LROL predictions stay closer to the 45 degree line than the ROL predictions do when the MNL predictions deviate likewise in Figure 5. This change illustrates the benefit of mitigated coefficient attenuation from modelling preference heterogeneity. Two explanations can be offered as to why the more flexible LHROL does not improve upon the LMNL predictions noticeably. First, the LMNL parameters have been estimated so that the model fits the first pseudo-choice data the best. Second, since only one fewer class is identified for LMNL and the consequentially blurred distinction is primarily about the extent of systematic trade-offs among many attributes, the loss of flexibility does not influence the weighted average predictions much.

Subsection 3.2 suggests that the form of joint heterogeneity in $\boldsymbol{\beta}$ and $\sigma$ found in the present application can be generally expected in other applications too. As a concluding remark, we discuss a convenient approach to impose and implement it within a continuous mixture model.

Fiebig *et al.* (2010) formulate the Generalised MNL (GMNL) model that directly extends mixed MNL models which specify $\boldsymbol{\beta}$ as multivariate normal random parameters. To this end, a scalar random parameter $\lambda$ is introduced, and the systematic component of utility is written as $\lambda\boldsymbol{\beta} \cdot \mathbf{x}_{njt}$.[18] $\lambda$ serves several purposes, one of which is operationalising interpersonal variation in the overall scale of utility. It is specified as a log-normal random variable independent of $\boldsymbol{\beta}$: $\log \lambda \sim N(-\tau^2/2, \tau^2)$ where $\tau$ is a parameter to be estimated.

Now suppose that GMNL is augmented by a *non-random* rank-specific scale parameter $\sigma$ as in HROL or C-LHROL, and fitted to rank-ordered data. Then, for each draw of $\lambda$, the effective scale in the second rank becomes $\lambda\sigma$. Even though $\sigma$ itself is non-random, the resulting model effectively exhibits interpersonal variation in the rank-specific scale through $\lambda$; in fact, the effect is equivalent to introducing a particular form of heterogeneity in $\sigma$ directly. Moreover, the effective rank-specific scale is necessarily larger when $\lambda$ is larger or the preferences are more deterministic.

$\sigma$-augmented GMNL would be appropriate for rank-ordered data analysis, also because random scaling by $\lambda$ allows approximating various patterns of preference heterogeneity that usual mixed logit models fail to capture (Fiebig *et al.*, 2010; Keane and Wasi, 2012). With such flexibility, the researcher can assume more safely that the underlying form of misspecification is purely related to the iid disturbance term. Doiron *et al.* (2011) have implemented this modelling approach, though their conceptual motivation is quite different and more straightforward; they interpret the resulting model as following from a GMNL version of the response construction process (4).

# 5 Discussion

The problem of unstable coefficients in the rank-ordered logit model has been traditionally interpreted as a sign that survey respondents fail to provide reliable rank-ordered responses. This paper shows that it may originate from the inherent sensitivity of the model to stochastic misspecification instead. Even a minor departure from the postulated random utility distribution can induce the coefficients to become unstable across ranks. The problem thus can be expected in almost any empirical work, regardless of data reliability, when rank-ordered logit is employed as a tractable approximation to an unknown data generating process. As we discuss and partly demonstrate, flexibly

[18] Our description actually corresponds to GMNL-II, which is the easiest version of GMNL to summarise. See Fiebig *et al.* (2010) for other variations.

specified mixed rank-ordered logit and mixed heteroskedastic rank-ordered logit models would deliver analytically useful results nevertheless.

The use of stated preference experiments has become commonplace in environmental economics, health economics and other research areas characterised by the scarcity of readily available non-experimental data. A typical experiment asks each respondent to make a choice among several alternatives, though the effective sample size can be conveniently increased by asking her to rank the presented alternatives. The trade-offs between choice and ranking experiments deserve reassessment, to the extent that the problem of unstable coefficients has contributed to the perception that rank-ordered responses are less reliable. On a related issue, Caparros *et al.* (2008) collect both rank-ordered and choice data using exactly the same survey design, and find that the same estimates are obtained when the rank-ordered responses are recoded as choices.

This paper has not addressed misspecified rank-ordered logit analysis involving labelled alternatives. For example, in Caparros *et al.*, each choice set comprises two reforestation programs and the status quo, requiring an alternative-specific intercept distinguishing the last. Such intercept may not react to the misspecified disturbance in the same way as other coefficients, since the best fitting intercept can vary much more dramatically across pseudo-choice datasets, potentially changing in signs. While detailed implications are left for future research, we note that the rank-ordered logit model with changing decision protocols, implemented by Ben-Akiva *et al.* (1992), is but a heteroskedastic rank-ordered logit model in which alternative-specific intercepts are allowed to shift freely across ranks.

The primary message this paper (re)affirms is that an econometric analysis of rank-ordered data can proceed from the same microeconomic standpoint as that of choice data. The object to be modelled can be conceptualised as the behaviour of random utility maximisers, in terms of which applied economists are accustomed to think, instead of the actual cognitive process of ranking survey respondents, which is a much less familiar terrain. This message mirrors the standard textbook treatment of rank-ordered data modelling (pp.764-770, Ruud, 2000; pp.156-159, Train, 2009) but has not been duly attended in the empirical literature, possibly due to the intuitive appeal of psychological explanations for the problem of unstable coefficients. Berry *et al.* (2004) and Train (2008) may serve as a guide for future econometric research using rank-ordered data. In these studies, the extra information rankings provide is exploited to estimate models featuring richer economic behaviour, instead of survey response strategies.

## Acknowledgement

Helpful comments from Agne Suziedelyte, Arghya Ghosh, Carol Propper, Denise Doiron, Denzil G. Fiebig, Nada Wasi and Stephen Whelan are gratefully acknowledged. This work was supported under Australian Research Council's Discovery Projects funding scheme (project number: DP0881205). The author has no conflict of interest.

## References

Anderson S, Palma A, Thisse J. 1992. *Discrete choice theory of product differentiation.* MIT Press: Cambridge, MA.

Beggs S, Cardell S, Hausman J. 1981. Assessing the potential demand for electric cars. *Journal of Econometrics* **16**: 1-19. DOI: 10.1016/0304-4076(81)90056-7

Ben-Akiva M, Morikawa T, Shiroishi F. 1992. Analysis of the reliability of preference ranking data. *Journal of Business Research* **24**: 149-164. DOI: 10.1016/0148-2963(91)90033-T

Berry S, Levinsohn J, Pakes A. 2004. Differentiated products demand system from a combination of micro and macro data. *Journal of Political Economy* **112**: 68-105. DOI: 10.1086/379939

Boyle K, Holmes T, Teisl M, Roe B. 2001. A Comparison of Conjoint Analysis Response Formats. *American Journal of Agricultural Economics* **83**: 441-54. DOI: 10.1111/0002-9092.00168

Calfee J, Winston C, Stempski R. 2001. Econometric issues in estimating consumer preferences from stated preference data: a case study of the value of automobile travel time. *Review of Economics and Statistics* **83**: 699-707. DOI: 10.1162/003465301753237777

Caparros A, Oviedo J, Campos P. 2008. Would you choose your preferred alternative? Comparing choice and recoded ranking experiments. *American Journal of Agricultural Economics* **90**: 843-855. DOI: 10.1111/j.1467-8276.2008.01137.x

Chapman R, Staelin R. 1982. Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research* **19**: 288-301.

Cramer J. 2007. Robustness of logit analysis: unobserved heterogeneity and misspecified disturbances. *Oxford Bulletin of Economics and Statistics* **69**: 545-555. DOI: 10.1111/j.1468-0084.2007.00445.x

Daksvik J, Liu G. 2009. A framework for analyzing rank-ordered data with application to automobile demand. *Transportation Research Part A* **43**: 1-12. DOI: 10.1016/j.tra.2008.06.005

Doiron D, Hall J, Kenny P, Street D. 2011. Job preferences of students and new graduates in nursing. *CHERE Working Paper 2011/2*.

Fiebig D, Keane M, Louviere J, Wasi N. 2010. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* **29**: 393-421. DOI: 10.1287/mksc.1090.0508

Fok D, Paap R, Van Dijk B. 2012. A rank-ordered logit model with unobserved heterogeneity in ranking capabilities. *Journal of Applied Econometrics* **27**: 831-846. DOI: 10.1002/jae.1223

Foster V, Mourato G. 2002. Testing for consistency in contingent ranking experiments. *Journal of Environmental Economics and Management* **44**: 309-328. DOI: 10.1006/jeem.2001.1203

Greene W. 2008. *Econometric Analysis*, 6th ed. Pearson Prentice Hall: New Jersey.

Hausman J, Ruud P. 1987. Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics* **34**: 83-104. DOI: 10.1016/0304-4076(87)90068-6

Hess S, Train K. 2011. Recovery of inter- and intra-personal heterogeneity using mixed logit models. *Transportation Research Part B* **45**: 973-990. DOI: 10.1016/j.trb.2011.05.002

Hole A. 2007. Fitting mixed logit models by using maximum simulated likelihood. *Stata Journal* **7**: 388-401.

Huber J, Train K. 2001. On the similarity of classical and Bayesian estimates of individual mean partworths. *Marketing Letters* **12**: 259-269. DOI: 10.1023/A:1011120928698

Keane M, Wasi N. 2012. Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics*: in press. DOI: 10.1002/jae.2304

Kenny P, Doiron D, Hall J, Street D, Milton-Wildey K, Parmenter G. 2012. The training and job decisions of nurses - the first year of a longitudinal study investigating nurse recruitment and retention. *CHERE Working Paper 2012/02.*

Layton D. 2000. Random coefficient models for stated preference surveys. *Journal of Environmental Economics and Management* **40**: 21-36. DOI: 10.1006/jeem.1999.1104

Luce R. 1959. *Individual choice behavior: a theoretical analysis.* Wiley: New York.

McFadden D. 1981. Econometric models of probabilistic choice. In *Structural analysis of discrete data with econometric applications*, Manski C, McFadden D (ed): 198-272. MIT Press: Cambridge, MA.

McFadden D, Train K. 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics* **15**: 447-470. DOI: 10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1

Naude M, McCabe R. 2005. Magnet hospital research pilot project conducted in hospitals in Western Australia. *Contemporary Nurse* **20**: 38–55. DOI: 10.5172/conu.20.1.38

Ruud P. 2000. *An introduction to classical econometric theory.* Oxford University Press: New York.

Scarpa R, Notaro S, Louviere J, Raffaeli R. 2011. Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in Alpine Grazing Commons. *American Journal of Agricultural Economics* **93**: 813-828. DOI: 10.1093/ajae/aaq174

Seago J, Ash M, Spetz J, Coffman J, Grumbach K. 2001. Hospital registered nurse shortages: Environmental, patient, and institutional predictors. *Health Services Research* **36**: 831-852.

Siikamaki J, Layton D. 2007. Discrete choice survey experiments: a comparison using flexible methods. *Journal of Environmental Economics and Management*, **53**: 122-139. DOI: 10.1016/j.jeem.2006.04.003

Train K. 2008. EM Algorithms for Nonparametric Estimation of Mixing Distributions. *Journal of Choice Modelling* **1**: 40-69.

Train K. 2009. *Discrete choice methods with simulation*, 2nd ed. Cambridge University Press: New York.

Train K, Winston C. 2007. Vehicle choice behavior and the declining market share of U.S. automakers. *International Economic Review* **48**: 1469-1496. DOI: 10.1111/j.1468-2354.2007.00471.x

Vossler C, Doyon M, Rondeau D. 2012. Truth in consequentiality: theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics* **4**: 145-171. DOI: 10.1257/mic.4.4.145

Walker J, Ben-Akiva M, Bolduc D. 2007. Identification of parameters in normal error component logit-mixture (NECLM) models. *Journal of Applied Econometrics* **22**: 1095-1125. DOI: 10.1002/jae.971

Yoo H, Doiron D. 2012. The use of alternative preference elicitation methods in complex discrete choice experiments. *UNSW Australian School of Business Research Paper No. 2012-16.*

# Appendices

## Appendix 1. A summary of main simulated examples

Each table in this appendix reports the mean (outside the brackets) and the range (inside the brackets) of each parameter estimate over 100 artificial datasets. Column DGP lists the true utility weights on the observed attributes that have been used to simulate the utility of each alternative. Column $Q$=$q$ summarises the ROL estimates obtained by using the response variable detailing the top $q$ ranks.

Table 4: Example 1 - summarised estimation results

| | DGP | ROL | | | | HROL |
|---|---|---|---|---|---|---|
| | | *Q=1* | *Q=2* | *Q=3* | *Q=4* | |
| $x_1$ | 1 | 1.109<br>[1.019,1.179] | 1.001<br>[0.946,1.047] | 0.934<br>[0.892,0.970] | 0.889<br>[0.847,0.930] | 1.110<br>[1.047,1.185] |
| $x_2$ | 1 | 1.108<br>[1.046,1.177] | 0.998<br>[0.948,1.049] | 0.930<br>[0.890,0.962] | 0.887<br>[0.851,0.925] | 1.106<br>[1.021,1.171] |
| $\sigma_2$ | | | | | | 0.798<br>[0.741,0.881] |
| $\sigma_3$ | | | | | | 0.687<br>[0.634,0.779] |
| $\sigma_4$ | | | | | | 0.589<br>[0.521,0.657] |

Table 5: Example 2 - summarised estimation results

| | DGP | ROL | | | HROL |
|---|---|---|---|---|---|
| | | *Q=1* | *Q=2* | *Q=3* | |
| price | -0.635 | -0.664<br>[-0.789,-0.543] | -0.598<br>[-0.709,-0.505] | -0.563<br>[-0.628,-0.473] | -0.664<br>[-0.780,-0.547] |
| contract | -0.140 | -0.155<br>[-0.202,-0.115] | -0.135<br>[-0.165,-0.107] | -0.122<br>[-0.152,-0.099] | -0.155<br>[-0.188,-0.127] |
| local | 1.431 | 1.556<br>[1.306,1.749] | 1.363<br>[1.182,1.507] | 1.250<br>[1.113,1.385] | 1.561<br>[1.316,1.771] |
| wknown | 1.055 | 1.149<br>[0.877,1.331] | 1.008<br>[0.860,1.192] | 0.908<br>[0.795,1.037] | 1.162<br>[0.961,1.374] |
| tod | -5.699 | -5.971<br>[-7.160,-4.838] | -5.360<br>[-6.211,-4.621] | -5.040<br>[-5.634,-4.323] | -5.957<br>[-6.983,-4.928] |
| seasonal | -5.900 | -6.171<br>[-7.249,-5.080] | -5.550<br>[-6.395,-4.740] | -5.210<br>[-5.811,-4.424] | -6.174<br>[-7.268,-5.138] |
| $\sigma_2$ | | | | | 0.740<br>[0.638,0.874] |
| $\sigma_3$ | | | | | 0.589<br>[0.421,0.780] |

Table 6: Example 3 - summarised estimation results

| | DGP | ROL | | HROL |
|---|---|---|---|---|
| | | *Q=1* | *Q=2* | |
| public hosp | 0.239 | 0.247 | 0.200 | 2.430 |
| | | [0.154,0.348] | [0.126,0.282] | [0.158,0.333] |
| 3 rotations | 0.212 | 0.221 | 0.206 | 0.220 |
| | | [0.130,0.287] | [0.138,0.260] | [0.145,0.278] |
| flex hours | 0.138 | 0.149 | 0.125 | 0.151 |
| | | [0.048,0.244] | [0.060,0.197] | [0.074,0.228] |
| flex rost | 0.568 | 0.588 | 0.512 | 0.590 |
| | | [0.463,0.688] | [0.432,0.594] | [0.496,0.692] |
| well staff | 0.409 | 0.410 | 0.381 | 0.421 |
| | | [0.312,0.540] | [0.307,0.489] | [0.344,0.539] |
| supp mgt | 1.032 | 1.049 | 0.938 | 1.050 |
| | | [0.937,1.156] | [0.865,1.026] | [0.967,1.142] |
| well equip | 0.375 | 0.373 | 0.350 | 0.376 |
| | | [0.241,0.490] | [0.244,0.453] | [0.257,0.493] |
| encourage | 0.545 | 0.568 | 0.496 | 0.572 |
| | | [0.484,0.656] | [0.419,0.568] | [0.489,0.641] |
| abund park | 0.081 | 0.094 | 0.067 | 0.087 |
| | | [0.000,0.192] | [-0.003,0.144] | [0.005,0.166] |
| app resp | 0.462 | 0.469 | 0.410 | 0.461 |
| | | [0.372,0.570] | [0.340,0.489] | [0.379,0.557] |
| excell care | 0.822 | 0.829 | 0.743 | 0.825 |
| | | [0.746,0.950] | [0.677,0.838] | [0.754,0.930] |
| log salary | 2.807 | 2.887 | 2.573 | 2.894 |
| | | [2.644,3.292] | [2.373,2.820] | [2.598,3.154] |
| $\sigma_2$ | | | | 0.735 |
| | | | | [0.665,0.823] |

# Appendix 2. Detailed estimation results - latent class models

Table 7: Detailed estimation results - latent class models

| | LMNL | | | LROL | | | |
|---|---|---|---|---|---|---|---|
| | Class 1 | Class 2 | Class 3 | Class 1 | Class 2 | Class 3 | Class 4 |
| public hosp | 0.306** | 0.245* | 0.294 | 0.079 | 0.006 | 0.293 | 0.205** |
| | (0.066) | (0.123) | (0.185) | (0.087) | (0.114) | (0.343) | (0.044) |
| 3 rotations | 0.317** | 0.388** | 0.157 | 0.089 | 0.129 | 0.600 | 0.223** |
| | (0.067) | (0.139) | (0.175) | (0.086) | (0.117) | (0.392) | (0.043) |
| flex hours | 0.110 | 0.417** | 0.163* | 0.102 | 0.121 | 0.034 | 0.120** |
| | (0.057) | (0.139) | (0.168) | (0.085) | (0.112) | (0.120) | (0.043) |
| flex rost | 0.616** | 0.905** | 0.459** | 0.289** | 0.094 | 0.518** | 0.598** |
| | (0.069) | (0.157) | (0.186) | (0.096) | (0.135) | (0.120) | (0.046) |
| well staff | 0.384** | 0.578** | 0.763** | 0.379** | 0.669** | 0.435** | 0.350** |
| | (0.059) | (0.127) | (0.202) | (0.088) | (0.123) | (0.139) | (0.044) |
| supp mgt | 1.404** | 0.575** | 1.004** | 0.776** | 0.585** | 3.536** | 0.738** |
| | (0.092) | (0.155) | (0.200) | (0.109) | (0.141) | (0.742) | (0.074) |
| well equip | 0.534** | 0.226 | 0.754** | 0.526** | 0.251* | 0.936* | 0.364** |
| | (0.073) | (0.135) | (0.190) | (0.092) | (0.119) | (0.405) | (0.044) |
| encourage | 0.629** | 0.613** | 0.413** | 0.366** | 0.361** | 1.239** | 0.487** |
| | (0.069) | (0.132) | (0.158) | (0.087) | (0.122) | (0.237) | (0.050) |
| abund park | 0.039 | 0.119 | 0.372* | 0.200* | 0.139 | 0.015* | 0.112** |
| | (0.058) | (0.120) | (0.174) | (0.089) | (0.125) | (0.118) | (0.045) |
| app resp | 0.599** | 0.382** | 0.642** | 0.562** | 0.400** | 1.098** | 0.413** |
| | (0.076) | (0.138) | (0.188) | (0.110) | (0.118) | (0.246) | (0.049) |
| excell care | 0.615** | 0.488** | 3.005** | 2.211** | 0.578** | 1.341** | 0.400** |
| | (0.091) | (0.133) | (0.337) | (0.157) | (0.108) | (0.301) | (0.057) |
| log salary | 1.324** | 8.356** | 1.834** | 1.481** | 9.766** | 0.886 | 1.737** |
| | (0.336) | (1.073) | (0.695) | (0.345) | (0.646) | (0.502) | (0.168) |
| $\pi_c$ | 0.577** | 0.247** | 0.175** | 0.204** | 0.141** | 0.140** | 0.515** |
| | (0.053) | (0.050) | (0.032) | (0.033) | (0.018) | (0.040) | (0.042) |
| log-likelihood | | -3315.88 | | | -5785.35 | | |
| bic | | 6869.84 | | | 11890.24 | | |
| parameters | | 38 | | | 51 | | |

*(Continued on next page)*

Table 7. (*Continued*)

| | C-LHROL | | | | LHROL | | | |
|---|---|---|---|---|---|---|---|---|
| | Class 1 | Class 2 | Class 3 | Class 4 | Class 1 | Class 2 | Class 3 | Class 4 |
| public hosp | 0.128 | 0.016 | 0.641** | 0.282** | 0.136 | 0.010 | 0.484 | 0.295** |
| | (0.096) | (0.146) | (0.359) | (0.052) | (0.097) | (0.125) | (0.362) | (0.054) |
| 3 rotations | 0.131 | 0.162 | 1.102** | 0.272** | 0.132 | 0.112 | 0.837* | 0.286** |
| | (0.099) | (0.150) | (0.378) | (0.052) | (0.101) | (0.128) | (0.405) | (0.055) |
| flex hours | 0.155 | 0.211 | 0.054 | 0.155** | 0.163 | 0.143 | 0.007 | 0.179** |
| | (0.096) | (0.140) | (0.152) | (0.050) | (0.098) | (0.124) | (0.134) | (0.052) |
| flex rost | 0.374** | 0.181 | 0.646** | 0.764** | 0.381** | 0.042 | 0.536** | 0.813** |
| | (0.103) | (0.163) | (0.170) | (0.056) | (0.104) | (0.150) | (0.161) | (0.059) |
| well staff | 0.471** | 0.799** | 0.661** | 0.414** | 0.479** | 0.693** | 0.535** | 0.426** |
| | (0.100) | (0.152) | (0.184) | (0.052) | (0.102) | (0.138) | (0.176) | (0.054) |
| supp mgt | 1.041** | 0.751** | 5.088** | 0.913** | 1.060** | 0.624** | 4.334** | 0.939** |
| | (0.116) | (0.186) | (0.770) | (0.074) | (0.121) | (0.161) | (0.864) | (0.074) |
| well equip | 0.614** | 0.270 | 1.619** | 0.416** | 0.625** | 0.266* | 1.269** | 0.425** |
| | (0.101) | (0.155) | (0.403) | (0.053) | (0.104) | (0.134) | (0.460) | (0.054) |
| encourage | 0.480** | 0.450** | 1.790** | 0.588** | 0.493** | 0.387** | 1.550** | 0.603** |
| | (0.097) | (0.152) | (0.399) | (0.057) | (0.099) | (0.136) | (0.372) | (0.059) |
| abund park | 0.230* | 0.271 | -0.066 | 0.129* | 0.221* | 0.198 | -0.024 | 0.142** |
| | (0.094) | (0.161) | (0.152) | (0.051) | (0.095) | (0.139) | (0.133) | (0.053) |
| app resp | 0.818** | 0.445** | 1.174** | 0.465** | 0.841** | 0.417** | 1.083** | 0.461** |
| | (0.112) | (0.153) | (0.339) | (0.058) | (0.114) | (0.132) | (0.285) | (0.059) |
| excell care | 2.472** | 0.658** | 1.929** | 0.434** | 2.493** | 0.578** | 1.614** | 0.456** |
| | (0.154) | (0.136) | (0.494) | (0.064) | (0.169) | (0.119) | (0.439) | (0.066) |
| log salary | 1.245** | 11.875** | 1.412* | 2.324** | 1.120** | 10.553** | 1.018 | 2.558** |
| | (0.343) | (0.956) | (0.693) | (0.215) | (0.344) | (0.824) | (0.619) | (0.224) |
| $\pi_c$ | 0.235** | 0.137** | 0.113** | 0.514** | 0.237** | 0.131** | 0.117** | 0.516** |
| | (0.030) | (0.019) | (0.023) | (0.035) | (0.030) | (0.017) | (0.025) | (0.035) |
| $\sigma_c$ | 0.615† | | | | 0.585† | 0.923 | 0.816 | 0.514† |
| | (0.032) | | | | (0.059) | (0.115) | (0.147) | (0.045) |
| log-likelihood | -5738.36 | | | | -5730.03 | | | |
| bic | 11802.50 | | | | 11804.64 | | | |
| parameters | 52 | | | | 55 | | | |

Standard errors in parentheses have been adjusted for clustering at the respondent level. $*$ significant at the 5% level; $**$ significant at the 1% level; † significantly less than the unity at the 1% level.