A Model for Ordered Data with Clustering of Observations*

Tim R. L. Fry
Department of Econometrics and Business Statistics
Monash University

and

Mark N. Harris
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne

Melbourne Institute Working Paper No. 2/00

ISSN 1328-4991

ISBN 0 7340 1482 1

January 2000

* We are grateful to Adrian Pagan and attendees of a Melbourne Institute workshop for useful comments and to Dave Fournier for computing advice. The usual caveats apply. The first author kindly acknowledges funding from a Monash University Faculty Grant. The second, support from an Australian Research Council grant.

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* (03) 9344 5330**
***Fax* (03) 9344 5630**
***Email* melb.inst@iaesr.unimelb.edu.au**
***WWW Address* http://www.ecom.unimelb.edu.au/iaesrwww/home.html**

A Model for Ordered Data with Clustering of Observations*

Tim R.L. Fry[a] and Mark N. Harris[b]

[a]Department of Econometrics and Business Statistics,
Monash University,
Clayton, Victoria 3800, Australia.
Email: Tim.Fry@BusEco.monash.edu.au

[b]Institute of Applied Economic and Social Research
Melbourne University
Parkville, Victoria 3052, Australia
Email: mark.harris@iaesr.unimelb.edu.au

Abstract: At present, there appears to be no qualitative dependent model that can simultaneously account for data sets in which the variable of interest has both a multi-modal distribution and is potentially ordered. Such a multi-modal distribution may be the result of individuals being captive to particular choices. Such a case arises when there is *digit preferencing* (particular numbers, such as 0, 5 and 10, are often favored in many survey-based data sets). This paper introduces a new discrete choice model, the Dogit Ordered Extreme Value (DOGEV), that does account for both ordering and digit preferencing in the data, and applies it to an Australian Inflationary Expectations data set.

JEL Classification: C25, D12, E31.
Keywords: Ordered data, digit preferences, inflationary expectations.

Corresponding Author: Dr. Tim R.L. Fry, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia. Email: Tim.Fry@BusEco.monash.edu.au. Fax: IDD + 61 3 9905 5474. Voice: IDD + 61 3 9905 2415

1 Introduction

The use of qualitative response discrete choice models is now extremely widespread (for example, in the fields of economics, transportation research, marketing and so on). Predominantly this takes the form of analyzing unit data, often responses to survey questionnaires. A distinguishing feature of many such data sets, is that often there will be a clustering around particular responses. This is particularly the case where a numerical response is elicited. That is, the researcher is likely to observe *digit preferencing* - certain numbers, such as 0, 5 and 10 (presumably as a result of the decimal system) tend to be favoured. Moreover, again especially in instances where a numerical answer is given, there is also likely to be ordering in the observed responses.

Traditionally such models would be estimated using a multinomial logit (MNL) model or, if indeed one were concerned about any potential ordering in the data, ordered probit or logit models. More recently, extensions have been proposed to McFadden's (McFadden (1978)) *Generalized Extreme Value* class of models by Small (1987) and Small (1994). Here the MNL model is generalized so that alternatives are implicitly ordered and a stochastic correlation between choices of close proximity is introduced, yielding the *Ordered Extreme Value* (OGEV) model. However, none of these approaches specifically accounts for clustering of observed responses around several alternatives, as would be the case if digit preferencing were present in the data.

A model that can account for such a multi-modal distribution, is the DOGIT model (Gaudry and Dagenais (1979)). The DOGIT model contains additional parameters to the standard MNL model, which are most easily interpreted as "captivity", "loyalty", "gravity" or "preference" coefficients. That is, the DOGIT models allows individuals to be captive to particular choices, for example if the alternatives themselves possess qualities that make them more attractive in some sense. Indeed, it is this particular aspect of the DOGIT model that makes it so appropriate for digit preferencing, as individuals tend to gravitate to integers in multiples of 5 and 10, presumably for psychological reasons.

In spite of its attractiveness in many situations, the DOGIT model has been rarely applied in practice (for example, see Bordley (1990), Chandrasekharan *et al* (1994), Gaudry (1980), Gaudry and Wills (1979), and Tse (1987)). A potential drawback of the DOGIT model however, is that, as with the MNL model, it does not allow for any ordering in the data. In this paper we introduce the Dogit Ordered Extreme Value (DOGEV) model, that combines the flexibility of the MNL model in allowing coefficients to vary across alternatives, with the ordering and correlation of proximate choice properties of the OGEV model, with the captivity properties of the DOGIT model.

We apply the DOGEV model to Australian inflationary expectations' data. From a policy perspective, it is important to understand what drives inflationary expectations. Such expectations have far reaching implications for the macroeconomy as a whole, especially in an environment of an active inflationary setting policy regime. Inflationary expectations are likely to affect the processes of wage bargaining, price setting and asset allocation, amongst other things, and as a consequence are likely to direct affect monetary policy and the activities of central banks. Moreover, this data appears particularly pertinent for such an application, as the numerical answers would appear to be ordered, and there is clear evidence of digit preferencing at such inflationary expectations of 0, 5 and 10%.

The plan of this paper is as follows. In Section 2 the DOGEV model is derived and the economic model described in Section 3. Section 4 describes the data and Section 5 the empirical results. Finally, Section 6 concludes.

2 An Ordered Probability Model with Digit Preferencing

As is common in the literature (Fry *et al* (1993)) we start with a random utility maximization model, with indirect utility function given by:

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \mathbf{x}_i' \boldsymbol{\beta}_j + \varepsilon_{ij}, \tag{1}$$

with $i = 1, \ldots N$, $j = 1, \ldots J$. U_{ij} is the utility individual i gains from alternative j, which is assumed to be a (linear) function of their $(k \times 1)$ vector of observed individual characteristics $\mathbf{x}_i$, with associated unknown weights $\boldsymbol{\beta}_j$ (that is, the coefficient vector varies across alternative). Finally, ε_{ij} is a random disturbance term assumed to follow an independent Type 1 Extreme Value distribution. The individual is assumed to choose the alternative which maximizes their utility. With choice set $C = \{1, \ldots, J\}$,the associated MNL probabilities are given by:

$$P_{ij}^{MNL} = \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})}, \tag{2}$$

where to identify the model the restriction that $\boldsymbol{\beta}_1 = \mathbf{0}$ is used (see Maddala (1989)).

2.1 The DOGIT Model

MNL models are extremely popular in practice. This is primarily due to their simplicity and ease of estimation. However, this simplicity does impose some very strong restrictions on the model, most notably that of the *Independence of Irrelevant Alternatives* (IIA).

The IIA property of MNL models, essentially says that the *odds ratio*, P_{ij}/P_{ik}, $j \neq k$, is independent of all other alternatives, and moreover independent of additions to, and deletions from, the full choice set. In many instances this appears to be an unrealistic assumption, a problem exacerbated by the fact that tests for IIA generally have very poor power properties (see Fry and Harris (1996)).

A number of non-IIA alternatives to the MNL have been proposed. The DOGIT model of Gaudry and Dagenais (1979) in particular appears attractive for many modelling instances, as its probabilities expand on the MNL ones of equation (2) so avoiding the IIA property. The DOGIT probabilities are given by

$$P_{ij}^{DOGIT} = \frac{\exp(V_{ij}) + \theta_j \sum_{k=1}^{J} \exp(V_{ik})}{\left(1 + \sum_{k=1}^{J} \theta_k\right) \sum_{k=1}^{J} \exp(V_{ik})}. \tag{3}$$

A requirement of the DOGIT model is that, to ensure a proper probability density function, the θ parameters, are non-negative, $\theta_j \geq 0 \ \forall \ j = 1, \ldots, J$.

The DOGIT model can also be conceptualized as arising from a two step choice process (Fry and Harris (1996)). In this interpretation an individual is either "captive" to one of the J outcomes or chooses freely from the full choice set. The selection probability is then decomposed into the sum of the probability of choosing outcome j as a captive choice plus the probability of exercising free choice multiplied by the probability of choosing outcome j freely from the full choice set. Given a suitable parameterization (see Fry and Harris *op cit* pp 24-25) the model is

$$P_{ij}^{DOGIT} = \frac{\theta_j}{1 + \sum_{k=1}^{J} \theta_k} + \frac{1}{1 + \sum_{k=1}^{J} \theta_k} \times P_{ij}^{MNL}. \tag{4}$$

The parameterization of equation (4) illustrates a further boundary condition on the admissable range for the θ_j values (in addition to $\theta_j \geq 0 \ \forall \ j = 1, \ldots, J$). In the limit, the proportion choosing alternative j in a sample, must be greater or equal to the proportion given by the "captive" probability $P_{ij}^{captive}$, where

$$P_{ij}^{captive} = \frac{\theta_j}{1 + \sum_{k=1}^{J} \theta_k}.$$

Effectively, this places an upper bound on the admissable θ_j values.

In such a parameterization, the θ's can be interpreted as "preference", "loyalty" or "gravity" parameters. This interpretation is of particular interest in our application as the preference idea can be viewed as analogous to digit preferencing, or alternatively that individuals are drawn to particular numbers.

In the context of inflationary expectations, the DOGIT model assumes that people simply prefer certain numbers. The fact that *why* they do so, is immaterial.It is possible to generalise the DOGIT model further by allowing these gravity parameters to be a function of observed heterogeneity. However, this is not considered in this paper, as the DOGEV model is already deemed to be sufficiently heavily parameterised. At one extreme, if the pull of these gravity parameters is "large" for any particular alternative they are likely to dominate the ultimate choice probabilities for that alternative - irrespective of observed personal heterogeneity. At the other extreme, a zero θ value for an alternative, will result in choice probabilities being driven solely by observed heterogeneity. In between these extremes, choice probabilities will be a combination of the two. An example might be that if headline inflation has been at 2% for several periods, $\theta_{j=2}$ may be such that the probability of choosing 2% is predominantly unaffected by individuals' characteristics which otherwise one would have expected to be influential. Indeed, individuals are simply drawn to this alternative. Essentially, the θ's are effectively controlling for unobserved heterogeneity, not of the individual, but of the choice alternative.

2.2 The OGEV Model

Small (1987) introduces a discrete choice model, the Ordered Generalized Extreme Value (OGEV), that is a member of the Generalized Extreme Value class of models. Again, the OGEV probabilities expand on the MNL ones of equation (2) such that IIA is no longer embodied. However, the underlying motivation for the OGEV model is to provide a suitable model for alternatives that are ordered in some sense. Unlike the MNL or DOGIT probabilities, the OGEV ones embody a correlation between alternatives in close proximity.

Such a correlation appears likely for ordered data in many instances, especially where the observed outcomes are realizations of an underlying latent scale. For example, given a five-point response scale ($j = 1, \ldots, 5$) say, of satisfaction, individuals may choose "neutral" ($j = 3$), but be influenced by the neighboring choices of "moderately satisfied" ($j = 4$) and "moderately dissatisfied" ($j = 2$). Or, in terms of inflationary expectations, one can naturally expect a correlation between choices say of $j = 2\%$ and $j = 3\%$ - especially if the "true" expectation is 2.5%. Moreover, apparent digit preferencing may be the result of individuals rounding from wide notional bands - a price expectation of anywhere between 2.5% and 7.5% may yield an observed response of 5%. Again, if this were the case, it would be natural to expect a correlation between the observed choice and neighbouring ones.

Although it is possible to allow the window of correlation to be arbitrarily large, this increases the number of parameters to be estimated and makes estimation cumbersome (Small (1987)). Therefore we restrict attention to the standard OGEV model (Small *op cit* p414). In Smalls' notation we have $M = 2$ and $\rho_r = \rho$ for all r. However, with regard to inflationary expectations, it is possible that the more flexible correlation structure implied by multiple ρ might be more appropriate. For reasons of parismony we do not consider this model variant.

The standard OGEV model implies a correlation between outcomes that are near neighbors. Analogously to a moving average process, this correlation decreases the further away two outcomes j and k are, and moreover is zero when $|j - k| > 2$. Although they cannot be written explicitly in closed form, these correlations are *inversely* related to the parameter ρ. The standard

OGEV probabilities are given by

$$P_{ij}^{OGEV} = \frac{\exp\left(\rho^{-1}V_{ij}\right)}{\sum_{r=1}^{J+1}\left(\exp\left(\rho^{-1}V_{ir-1}\right)+\exp\left(\rho^{-1}V_{ir}\right)\right)^{\rho}} \tag{5}$$

$$\times \Big[\left(\exp\left(\rho^{-1}V_{ij-1}\right)+\exp\left(\rho^{-1}V_{ij}\right)\right)^{\rho-1}$$
$$+ \left(\exp\left(\rho^{-1}V_{ij}\right)+\exp\left(\rho^{-1}V_{ij+1}\right)\right)^{\rho-1}\Big],$$

with the convention that $\rho^{-1}V_{i0} = \rho^{-1}V_{iJ+1} = 0$ and $0 < \rho \leq 1$.

As $\rho \rightarrow 1$, OGEV probabilities converge to MNL ones, which gives a simple parameter restriction ($\rho = 1$) based test of the OGEV versus MNL formulations. Such a test is also implicitly a test of ordering versus non-ordering of the outcomes in the choice set. Finally, we note that as $\rho \rightarrow 0$, the associated cumulative distribution function is a degenerate one, but one still consistent with random utility maximization (Small (1987)).

2.3 The DOGEV Model

Unfortunately, the OGEV model does not allow for the phenomenon of digit preferencing and the DOGIT model does not allow either for the ordering of outcomes or for the (potential) correlation of proximate outcomes. In this section we combine the ideas underlying both the DOGIT and OGEV models to produce a new discrete choice model, the Dogit Ordered Extreme Value (DOGEV) model.

The idea is simple, combine the two step choice generating process of the DOGIT with the proximate correlation and ordering of the OGEV model. Specifically, we assume that the first step in the choice process for "captive", or digit preference, probabilities is the same as in the DOGIT formulation. That is, we allow the alternatives themselves to possess unobserved heterogeneity/characteristics which potentially attract individuals to them. This attraction is in addition to any the alternatives already possess, as determined by the individual and their observed characteristics.

In the second step, if free choice is exercised, the selection probabilities are given by the OGEV formulation. Thus, using equations (4) and (5) the selection probabilities for the DOGEV model are given by

$$P_{ij}^{DOGEV} = \frac{\theta_j}{1+\sum_{k=1}^{M}\theta_k} + \frac{1}{1+\sum_{k=1}^{M}\theta_k} \times P_{ij}^{OGEV}. \tag{6}$$

The nested models are therefore

$$\begin{array}{ll} \text{OGEV:} & \theta_1 = \ldots = \theta_M = 0,\ 0 < \rho \leq 1, \\ \text{DOGIT:} & \rho = 1, \text{ at least one } \theta_j > 0,\ j = 1, \ldots, J, \\ \text{MNL:} & \theta_1 = \ldots = \theta_M = 0,\ \rho = 1, \end{array}$$

of which the last two do not imply ordering in the observed outcomes. It should be noted that this model is particularly flexible as it allows us to model multi-modal ordinal data in a simple, but flexible way. Particular models are nested within DOGEV in a way that allows for simple hypothesis testing based model selection. Thus the DOGEV model is not restricted to applications where digit preferencing (or clustering of observed responses) and ordering is a potential data issue, but through its nested variants (DOGIT, OGEV and MNL) may be used in any application where the distribution of observed outcomes is either multi-modal or ordered or unordered.

The parameters of the DOGEV model can be consistently estimated using the maximum likelihood criterion. By defining an indicator variable d_{ij} as

$$d_{ij} = \begin{cases} 1 & \text{if individual } i \text{ chooses alternative } j \\ 0 & \text{otherwise} \end{cases}$$

the log-likelihood function will be

$$L^k(\phi\,|X) = \sum_{j=1}^{J}\sum_{i=1}^{N} d_{ij} \ln P_{ij}^k, \tag{7}$$

with k =MNL, DOGIT, OGEV and DOGEV and with $\phi' = \left[\text{vec}\beta_j\right]$, $\left[\left(\text{vec}\beta_j\right)', \theta'\right]$, $\left[\left(\text{vec}\beta_j\right)', \rho\right]$ and $\left[\left(\text{vec}\beta_j\right)', \theta', \rho\right]$, respectively.

3 An Application - Inflationary Expectations Data

Inflationary expectations have wide reaching implications for the economy as a whole - especially so in a policy regime of active inflation targeting. For example, The Reserve Bank of Australia is currently committed to an annual rate of inflation in the range of 2 - 3%. Such expectations are likely to affect price setting, wage bargaining and asset allocation, for example. For monetary authorities, therefore, it is important to know what drives inflationary expectations.

3.1 Previous Evidence

A significant amount of work has been undertaken modeling inflationary expectations in a time-series framework (for example, DeBrouwer and Ellis (1998) and Brischetto and DeBrouwer (1999)), although this is not of direct relevance here. However, presumably as a result of a lack of appropriate data sources, very little work appears to have been undertaken which seeks to explain inflationary expectations at an individual level. Exceptions to this are Harris and Harding (1998a), Harris and Harding (1998b) and Brischetto and DeBrouwer (1999). Moreover, these papers all use the same data set (although over differing periods) as utilized in this study.

Harris and Harding (1998a) considered low/medium/high inflationary expectations and estimated a sample selection-ordered probit model. However, in doing so they effectively side-stepped the important issue of digit preferencing. In a subsequent paper Harris and Harding (1998b), an attempt at modeling digit preferencing was undertaken by estimating a multi-stage, ordered probit-type model. However, this approach relies on the assumption that individuals observed to elicit an *a priori* determined digit preferenced outcome, were indeed "digit preferencers". In reality, those who are and those who simply had a genuine expectation which happened to coincide with a digit preference outcome, are observationally equivalent. Brischetto and DeBrouwer (1999) do not broach the issue of digit preferencing and estimate a simple ordinary least squares regression on a pooled number of survey months, spanning January 1995 to April 1998. In this way, any digit preferencing is ignored, as is the inherent discrete nature of the data. Moreover, over such an extended period, it would appear difficult to separate out the effects of personal demographics and headline inflation rates on expected inflation.

In the majority of theses studies, the demographics which appear to be important in determining inflationary expectations are invariably: age; education; household income; occupation and; gender.

3.2 Micro Determinants of Inflationary Expectations

Following Harris and Harding (1998a), Harris and Harding (1998b) and Brischetto and DeBrouwer (1999) (in part) we adopt a discrete choice framework within which it is assumed that in forming their inflationary expectations, individuals are utility maximizers. That is, the extent of "sophistication" of their forecasts (in terms of expenditures of time and resources in obtaining such) will be an increasing function of the benefits that they are likely to obtain from a more accurate forecast. Therefore, it is likely that, along with other (economic) variables, inflationary expectations will be driven by certain socio-demographic attributes that the individual possesses. In other words, different individuals will have access to different information sets. These different information sets will, in part, be a function of personal characteristics or attributes. Moreover, the extent to which these information sets are used will, again in part, be determined by the individual's utility maximizing process.

It is expected *a priori*, that the accumulation of knowledge with age is one variable amongst many others, that is likely to exert a positive influence on the sophistication of individuals' forecasts, as is the level of education. Those individuals with a close proximity to the price setting process, are likely to be more aware of the inflationary climate. In the empirical example, we choose occupations of "managers" and "sales-persons" to proxy these. An indicator of voting intentions is also included, as there is evidence that voters of the Opposition tend to have more pessimistic inflationary expectations (see MIAESR (1999) and Brischetto and DeBrouwer (1999)). We also include a dummy for place of residence, as this also may affect available information sets. It has been argued that there will be differences between male and female inflationary expectations, to the extent that one gender spends more time retail shopping than their counterparts (Bachelor and Jonung (1986)). Therefore, we also include a dummy variable for gender (again this can be interpreted as affecting information sets). These variables are summarized in Table 1.

INSERT TABLE 1 - About here

4 The Data

The data used are from the Melbourne Institute's Survey of Consumer Inflationary Expectations. The survey is a stratified random sample of 1,200 respondents, conducted monthly. Respondents are asked a wide variety of questions, including what they expect inflation to be over the coming year. In addition, a lot of personal demographics are also recorded. Although the survey is not a panel data set, to increase the sample size the most recent three months (February, March and April, 1999) were pooled and treated as a single cross section. As official headline inflation in Australia, the Consumer Price Index, is a quarterly series, and the fact that the three months chosen span a stable inflationary climate (the annual rate of increase in the CPI quarter 1 of 1999 was 1.2%), justifies such a pooling. More information on this series can be found in McDonnell (1994).

Observed inflationary expectations contained some quite extreme outliers, ranging from -50% to 80%. Estimation was based upon only those respondents who elicited a “sensible” expectation, defined as $0\% \leq P^e \leq 10\%$ (Brischetto and DeBrouwer (1999)). Of course, in a low inflation environment, some negative expected inflation rates could also be considered reasonable. However, for comparability with the previous work with this Australian data we adopt the same sample selection rule ($0\% \leq P^e \leq 10\%$). Selecting the data in this way yields results that are not based on unrealistic observations (outliers). Moreover, this procedure yields a data set with an appropriate number of alternatives, which additionally are contiguous. The (truncated) distribution of such expectations are shown in Figure 1.

INSERT FIGURE 1 - About here

We note that the annual rate of increase in Q1 of the C.P.I. was 1.2%. From Figure 1 it is evident, that *a priori*, there appears to be digit preferencing at expectations of 0, 5 and 10%. There is also significant mass at expectation points 2 and 3, which correspond to the recent headline rate, and moreover the Reserve Bank of Australia's inflation target range of 2-3%. The small number of missing values were removed, leaving a working sample of 2194 observations, down from an original sample of 2454.

In the unrestricted version of all of the above models, the response parameters $(\text{vec}\beta_j)$, vary across alternative. Consequently the number of parameters in $(\text{vec}\underline{\beta})$ where $\underline{\beta}$ is the stacked vector of β_j, $j = 1, \ldots, J$, is given by $(J-1)K$, where K is the number of explanatory variables including the constant term. So, in the interests of parsimony, the age, and income variables were entered as continuous (ordinal) variables (and not, for example, as a series of disaggregated indicator dummies).

The variables used in the study are summarized in Table 2.

INSERT TABLE 2 - About here

5 Empirical Results

5.1 Multinomial Logit

The estimation of a multinomial logit model is useful as a “benchmark” for two reasons. Firstly, we can see whether our chosen socio-economic variables are significant. Secondly, as the subsequent models in the (ordered) GEV class can prove somewhat troublesome in terms of estimation and convergence (see Small (1987) and Small (1994)), it is important to have the model well specified at this early stage. Table 3 contains the results for the multinomial logit model. The model appears to be well specified, with most of the variables generally significant. Note that the coefficients corresponding to the *a priori* digit preferenced outcomes (2, 5 and 10) tend to be the more significant ones. This may either be an effective sample size issue, or that the model “explains” these alternatives well.

INSERT TABLE 3 - About here

5.2 DOGEV

The parameter estimates for the DOGEV model are reported in Table 4. The model seems to be fairly well specified - as with the multinomial logit model earlier. Ordering appears to be important, as ρ is strongly significant. Indeed, ρ differs significantly from both zero and one. Since correlation is inversely related ρ to we note that the estimated value of 0.033 implies a high level of correlation between adjacent categories (expectations). Considering the captivity parameters we find significant parameters for expectations of 0%, 2%, 3%, 5% and 7% (θ_0, θ_2, θ_3, θ_5 and θ_7). Moreover, the largest effects are given by the modal choice, $j = 5\%$, with $j = 2\%$ and 3%, also exhibiting “large” effects. Thus we can say that these results clearly indicate that both captivity and ordering effects should be in our model. The DOGEV model is, therefore, clearly an improvement over all of its nested sub-models (MNL, OGEV and DOGIT).

INSERT TABLE 4 - About here

5.3 Some Model Evaluations

As stated in Section 2.3, the DOGEV model nests all of the MNL, DOGIT and OGEV models. Table 5 gives the results of likelihood ratio tests for the implicit parameter restrictions. Table 5 clearly illustrates the superiority of the DOGEV model over all nested variants.

INSERT TABLE 5 - About here

Using the above results, we now conduct a series of model evaluations (or predictions). Due to the complexity of such a model, it is unclear from the estimated coefficients how well the models describe the data, and moreover what its implications for the data are. We next present the results from a series of exercises which attempt to do this.

Firstly, in Figure 1 we present sample proportions of observed choices as a reference point, along with the implied extent of DOGEV "preference" for each expectation. In addition we also present the predicted probabilities of the DOGEV model evaluated at the sample means of the explanatory variables.

INSERT FIGURE 2 - About here

For choices with $\theta_j = 0$ (or small) there is no preferencing, and probabilities are driven by the OGEV probabilities. The DOGEV model appears well specified, attributing 60% of its probability for a typical individual for $j = 3$ to preferencing and 13% for $j = 10$.

In summary, allowing for ordering and captivity/preferencing in the data appears to yield sensible estimates. Moreover, in terms of predicted probabilities for a "typical" individual compared to sample proportions, the DOGEV model appears to fit the data extremely well, in addition to having a sensible interpretation.

We next consider the effect of particular explanatory variables on the DOGEV choice probabilities. Whether the individual was a Labor (Opposition) voter seemed to be persistently significant in the model estimation stages and Figure 3 illustrates the effect of this on the DOGEV choice probabilities. In this, and subsequent exercises, all *other* explanatory variables were evaluated at their sample means. The predicted probabilities closely mimic the sample proportions and we see that Labor voters are clearly more likely to have an expectation of 10% and less likely to have one of 0%. The other categories are relatively similar, except for $j = 3$, for which Labor voters are less likely to choose.

INSERT FIGURE 3 - About here

The results for selected education levels are given in Figure 4. Interestingly, individuals with either the highest or lowest educational level, are equally as likely to have an expectation of 0 and 5%. However, the educational effect of a tertiary qualification does appear to lower the probabilities of expectations of 6% and above.

INSERT FIGURE 4 - About here

Finally, we report the effects of gender on choice probabilities in Figure 5. Again, predictions generally closely follow sample proportions. It appears that males are relatively more likely to have lower inflationary expectations - having higher probabilities of choosing all rates between 0 and 4%, whilst females are more likely to higher expectations - generally having a higher probability of choosing any rate between 5 and 10%. Moreover, the digit preference effects of 5 and 10, appear to be significantly enhanced for females.

INSERT FIGURE 5 - About here

6 Conclusions

In this paper we set out to model Australian survey data on inflationary expectations, the distribution of which appeared to be both ordinal and multi-modal. We introduced a new model, the Dogit Ordered Extreme Value (DOGEV), that allows for both of these characteristics. The DOGEV model combines the flexibility of the multinomial logit (MNL) model in allowing coefficients to vary across alternatives, with the ordering and correlation of proximate outcomes properties of the ordered generalized extreme value (OGEV) model and with the captivity/preferencing/loyalty properties of the DOGIT model. It is a simple, parsimonious and flexible model that has the additional benefit of nesting certain key sub-models: the DOGIT, the OGEV and the MNL. Our application has shown that Australian inflationary expectations data are well modeled by the DOGEV with both significant captivity and ordering components. We believe that the new DOGEV model may be attractive in many other situations where one wishes to model unit record data that are both ordinal and multi-modal.

* **Acknowledgements:** We are grateful to Adrian Pagan and attendees of a Melbourne Institute workshop for useful comments, to Rosemary Doran for research assistance and to Dave Fournier for computing advice. The usual caveats apply. The first author kindly acknowledges funding from a Monash University Faculty Grant. The second, support from an Australian Research Council grant.

7 References

Bachelor, R. and L. Jonung (1986), "Information theory and group differences in inflation expectations", Discussion Paper 1, National Institute of Economic Research, Stockholm.

Bordley, R.F. (1990), "The Dogit model is applicable even without perfectly captive buyers", *Transportation Research - Series B*, **24B**, 315-323.

Brischetto, A. and G. DeBrouwer (1999), "Householders' inflation expectations", Discussion Paper 9903, Reserve Bank of Australia.

DeBrouwer, G. and L. Ellis (1998), "Forward-looking behaviour and credibility: Some evidence and implications for policy", Discussion Paper 9803, Reserve Bank of Australia.

Chandrasekharan, R., McCarthy, P.S. and G.P. Wright (1994), "Structural models of brand loyalty with an application to the automobile market", *Transportation Research - Series B*, **28B**, 445-462.

Fry, T.R.L., Brooks, R.D., Comley, B.R. and J. Zhang (1993), "Economic motivations for limited dependent and qualitative variable models", *Economic Record*, **69**, 193-205.

Fry, T.R.L. and M.N. Harris (1996), "A Monte Carlo study of tests for the independence of irrelevant alternatives property", *Transportation Research - Series B*, **30B**, 19-30.

Gaudry, M.J.I. (1980), "Dogit and Logit models of travel mode choice in Montreal", *Canadian Journal of Economics*, **13**, 268-279.

Gaudry, M.J.I. and M.G. Dagenais (1979), "The Dogit model", *Transportation Research - Series B*, **13B**, 105-112.

Gaudry, M.J.I. and M.J. Wills (1979), "Testing the Dogit model with aggregate time-series and cross-sectional travel data", *Transportation Research - Series B*, **13B**, 155-166.

Harris, M.N. and D. Harding (1998a), "An analysis of consumer inflationary expectations", mimeo Melbourne Institute of Applied Economic and Social Research.

Harris, M.N. and D. Harding (1998b), "A model for analysing survey data with digit preferencing", mimeo Melbourne Institute of Applied Economic and Social Research.

Maddala, G.S. (1989), *Limited dependent and qualitative variable models in econometrics*, Cambridge University Press: Cambridge.

McDonnell, J. (1994), "Measuring Australian consumer price expectations", Discussion Paper 7, Melbourne Institute of Applied Economic and Social Research.

McFadden, D. (1978), "Modeling the choice of residential location" in *Spatial Interaction Theory and Residential Location* (A. Karlqvist *et al*, eds.), North Holland: Amsterdam.

MIAESR (1999), "Survey of consumer inflationary expectations", Discussion Paper, Melbourne Institute of Applied Economic and Social Research.

Small, K. (1987), "A discrete choice model for ordered alternatives" *Econometrica*, **55**, 409-424.

Small, K. (1994), "Approximate generalized extreme value models of discrete choice", *Journal of Econometrics*, **62**, 351-382.

Tse, Y.K. (1987), "A diagnostic test for the multinomial Logit model", *Journal of Business & Economic Statistics*, **7**, 61-65.

Table 1: Micro Determinants of Inflationary Expectations

Variable	Description
Price Expectation	Numerical Answer to question: *How much do you think prices will increase/decrease?*
Gender	Indicator: 1 = Male
Age in Years	Bands: 18-24, 25-34, 35-44, 45-49, 50-54, 55-64, 65+
Income (Aus$000)	Bands: To 20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 100+
Place of Residence	Indicator: 1 = Urban
Occupation	Indicators: 1 = Manager; 1 = Salesperson
Education	Indicators: 1 = Up to Year 10; 1 = Tertiary or above
Tenancy	Indicators: 1 = Rented; 1 = Owned with Mortgage
Voting	Voted Labor (opposition) at last election

Table 2: Summary Statistics of Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Price Expectation	4.173	3.189	0	10
Male	0.538	0.499	0	1
Age	3.841	1.850	1	7
Income	4.440	2.918	1	10
Urban	0.6167	0.486	0	0
Manager	0.309	0.462	0	1
Sales	0.284	0.451	0	1
Low Education	0.259	0.438	0	1
Tertiary Education	0.428	0.495	0	1
Rented	0.189	0.391	0	1
Mortgage	0.360	0.480	0	1
Labor Voter	0.336	0.472	0	1

Table 3: Multinomial Logit Estimation Results

Constant	Male	Age	Income	Urban	Manager	Sales	Low Ed.	Uni. Ed.	Labor Voter	Rent	Mortgage
						j =1					
-0.267	-0.100	-0.161	-0.092	0.423	-0.210	-0.923	-0.087	0.259	0.406	-0.524	-0.800
(0.627)	(0.269)	(0.081)	(0.051)	(0.280)	(0.336)	(0.389)	(0.392)	(0.292)	(0.286)	(0.382)	(0.334)
						j = 2					
0.936	-0.304	-0.091	-0.068	-0.124	-0.170	-0.314	0.064	0.007	0.619	-0.087	0.047
(0.362)	(0.156)	(0.051)	(0.032)	(0.162)	(0.227)	(0.214)	(0.214)	(0.189)	(0.175)	(0.233)	(0.182)
						j = 3					
0.487	-0.183	-0.061	-0.079	-0.042	-0.110	-0.045	0.168	0.109	0.369	0.081	0.006
(0.383)	(0.162)	(0.054)	(0.031)	(0.167)	(0.238)	(0.220)	(0.225)	(0.194)	(0.185)	(0.244)	(0.205)
						j = 4					
-0.309	-0.125	-0.100	-0.110	-0.184	-0.313	-0.270	-0.152	0.433	0.866	-0.466	-0.072
(0.541)	(0.240)	(0.078)	(0.052)	(0.245)	(0.330)	(0.309)	(0.337)	(0.290)	(0.252)	(0.374)	(0.277)
						j = 5					
1.633	-0.615	-0.091	-0.101	0.074	-0.634	-0.302	0.111	0.025	0.709	0.067	-0.032
(0.341)	(0.143)	(0.048)	(0.029)	(0.150)	(0.209)	(0.190)	(0.194)	(0.175)	(0.162)	(0.216)	(0.174)
						j = 6					
-4.163	0.505	0.140	-0.145	0.357	0.057	0.123	0.116	-0.321	1.546	0.105	0.016
(1.248)	(0.603)	(0.177)	(0.114)	(0.549)	(0.530)	(0.722)	(0.689)	(0.654)	(0.561)	(0.795)	(0.610)
						j = 7					
-2.835	-0.502	0.124	-0.073	0.377	-0.397	-0.081	0.273	0.134	1.166	0.091	0.511
(0.970)	(0.350)	(0.124)	(0.073)	(0.374)	(0.518)	(0.481)	(0.451)	(0.468)	(0.342)	(0.586)	(0.409)
						j = 8					
-1.950	-0.188	0.040	-0.055	0.180	-0.212	-0.381	0.459	0.262	0.933	-0.425	-0.119
(0.711)	(0.291)	(0.099)	(0.057)	(0.314)	(0.411)	(0.400)	(0.397)	(0.361)	(0.304)	(0.478)	(0.346)

					j = 9						
0.454	0.095	-0.912	-0.118	-2.122	0.279	-0.268	0.898	-0.825	0.979	-6.720	-1.059
(1.242)	(0.609)	(0.364)	(0.141)	(1.053)	(0.793)	(0.935)	(0.949)	(1.156)	(0.722)	(0.501)	(0.790)
					j = 10						
1.619	-0.867	-0.143	-0.150	-0.149	-0.544	-0.480	0.270	0.083	1.196	-0.103	-0.092
(0.374)	(0.162)	(0.053)	(0.034)	(0.167)	(0.236)	(0.213)	(0.217)	(0.197)	(0.176)	(0.242)	(0.199)
log-L	-4192.376										

Standard Errors in parentheses.

Table 4: DOGEV Estimation Results

Constant	Male	Age	Income	Urban	Manager	Sales	Low Ed.	Uni. Ed.	Labor Voter	Rent	Mortgage
						j =1					
0.415	-0.062	-0.058	-0.127	0.024	-0.007	-0.664	-0.090	0.314	0.302	-0.386	-0.262
(0.201)	(0.082)	(0.024)	(0.030)	(0.074)	(0.082)	(0.236)	(0.121)	(0.120)	(0.120)	(0.154)	(0.113)
						j = 2					
-1.904	-0.127	-0.535	-0.104	-2.906	2.516	-0.871	-1.782	0.019	0.391	2.404	1.733
(40.829)	(0.262)	(0.192)	(0.051)	(11.348)	(1.113)	(0.823)	(3.956)	(0.301)	(0.293)	(40.827)	(40.828)
						j = 3					
0.964	-0.198	-0.085	-0.151	-0.135	0.144	0.007	-0.174	-0.011	0.518	-0.176	0.101
(0.441)	(0.188)	(0.057)	(0.038)	(0.175)	(0.288)	(0.287)	(0.248)	(0.229)	(0.230)	(0.264)	(0.209)
						j = 4					
1.065	-0.301	-0.076	-0.163	-0.174	-0.095	-0.002	-0.139	0.254	0.720	-0.242	0.111
(0.433)	(0.182)	(0.056)	(0.037)	(0.170)	(0.269)	(0.274)	(0.239)	(0.200)	(0.217)	(0.259)	(0.204)
						j = 5					
1.398	-0.719	-0.028	-0.155	0.059	-0.644	-0.126	0.302	0.360	0.859	0.155	0.173
(0.438)	(0.177)	(0.056)	(0.037)	(0.166)	(0.251)	(0.254)	(0.231)	(0.202)	(0.218)	(0.241)	(0.205)
						j = 6					
-4.733	0.588	0.367	-0.217	0.722	0.232	1.143	0.777	0.391	1.372	-0.043	-0.134
(1.912)	(0.698)	(0.218)	(0.130)	(0.578)	(0.884)	(0.670)	(0.599)	(0.853)	(0.445)	(0.647)	(0.642)
						j = 7					
-11.093	0.595	0.698	-0.318	2.716	-1.457	2.167	3.101	3.064	1.293	-0.170	-2.104
(6.391)	(0.519)	(0.331)	(0.195)	(3.707)	(70.710)	(1.051)	(3.716)	(3.772)	(0.472)	(0.605)	(4.688)
						j = 8					
-1.461	-0.224	0.048	-0.134	0.243	-0.011	-0.226	0.717	0.313	1.042	-0.469	0.016
(0.704)	(0.291)	(0.091)	(0.056)	(0.294)	(0.373)	(0.416)	(0.366)	(0.357)	(0.312)	(0.440)	(0.343)

						j = 9					
0.092	-0.090	-0.235	-0.184	-0.891	0.253	0.090	0.452	0.133	1.153	-2.175	-0.603
(0.704)	(0.344)	(0.107)	(0.061)	(0.481)	(0.395)	(0.508)	(0.405)	(0.385)	(0.359)	(5.081)	(0.418)
						j = 10					
1.979	-1.173	-0.131	-0.235	-0.113	-0.569	-0.377	0.441	0.279	1.558	-0.118	0.150
(0.469)	(0.230)	(0.061)	(0.051)	(0.202)	(0.311)	(0.330)	(0.283)	(0.249)	(0.266)	(0.262)	(0.246)
θ_0	0.080	(0.023)									
θ_1	0.000										
θ_2	0.255	(0.022)									
θ_3	0.127	(0.023)									
θ_4	0.000										
θ_5	0.118	(0.038)									
θ_6	0.001	(0.004)									
θ_7	0.024	(0.005)									
θ_8	0.000										
θ_9	0.000										
θ_{10}	0.034	(0.037)									
ρ	0.033	(0.006)									
log-L	-4164.869										

Standard Errors in parentheses.

Table 5: Likelihood Ratio Tests for DOGEV against Sub-Models

Sub-Model	Test Statistic Value	Degrees of Freedom
OGEV	54.993	11
DOGIT	30.919	1
MNL	55.014	12

Figure 1: Price Expectation (% per annum), N=2194

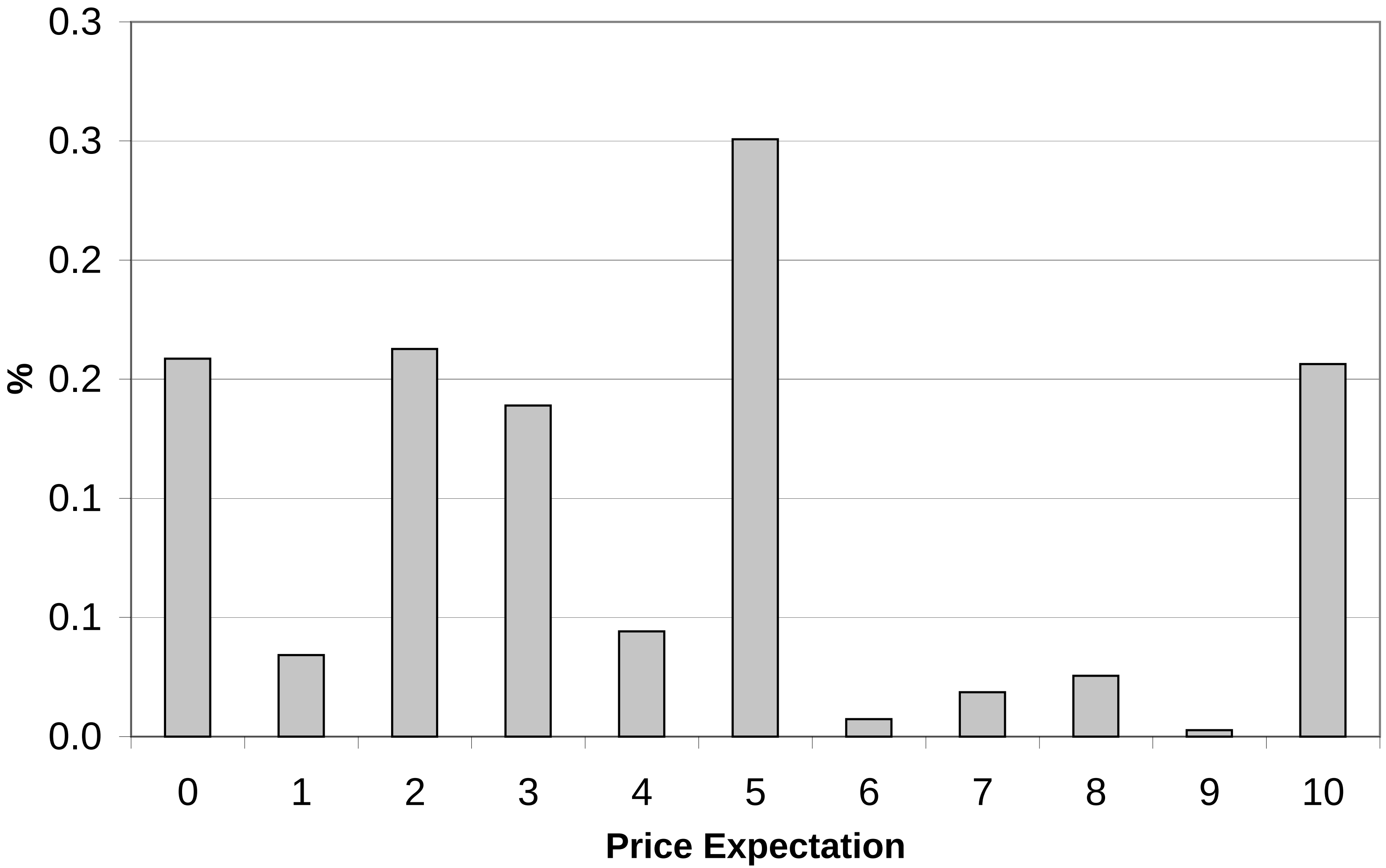

Figure 2: Sample Proportions and DOGEV Probabilities

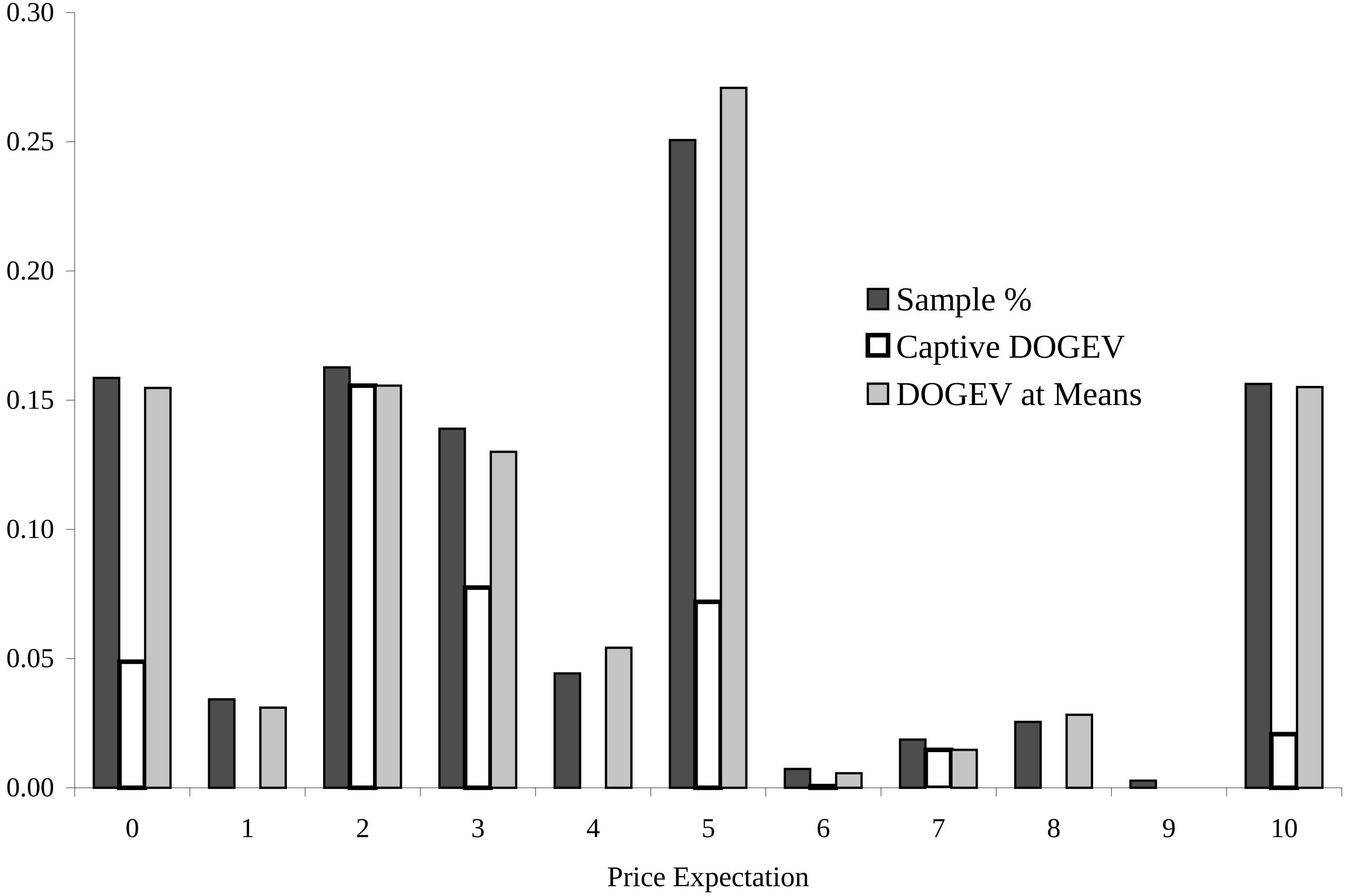

Figure 3: Fitted DOGEV Probabilities - Labor Voter

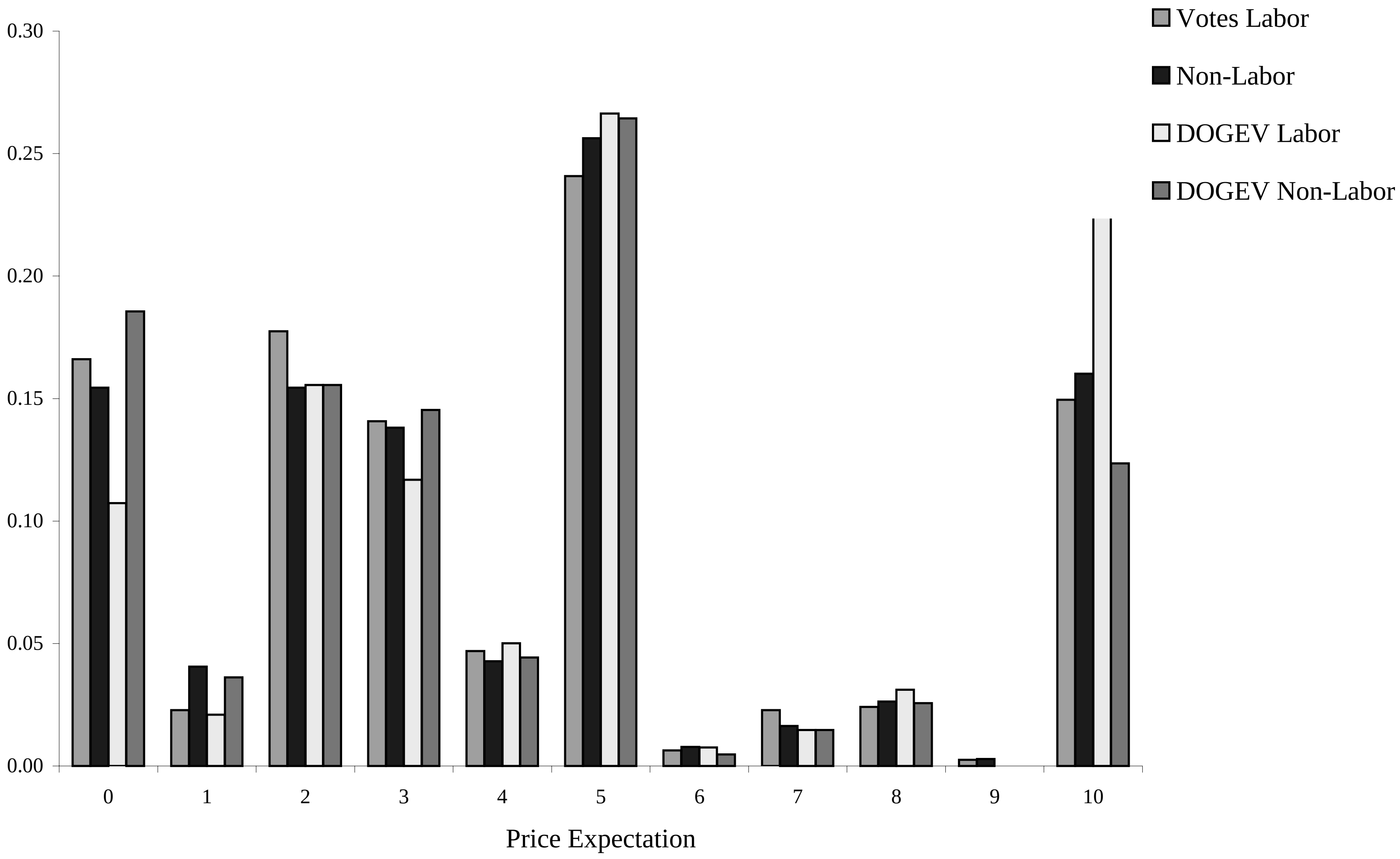

Figure 4: Fitted DOGEV Probabilities - Educational Level

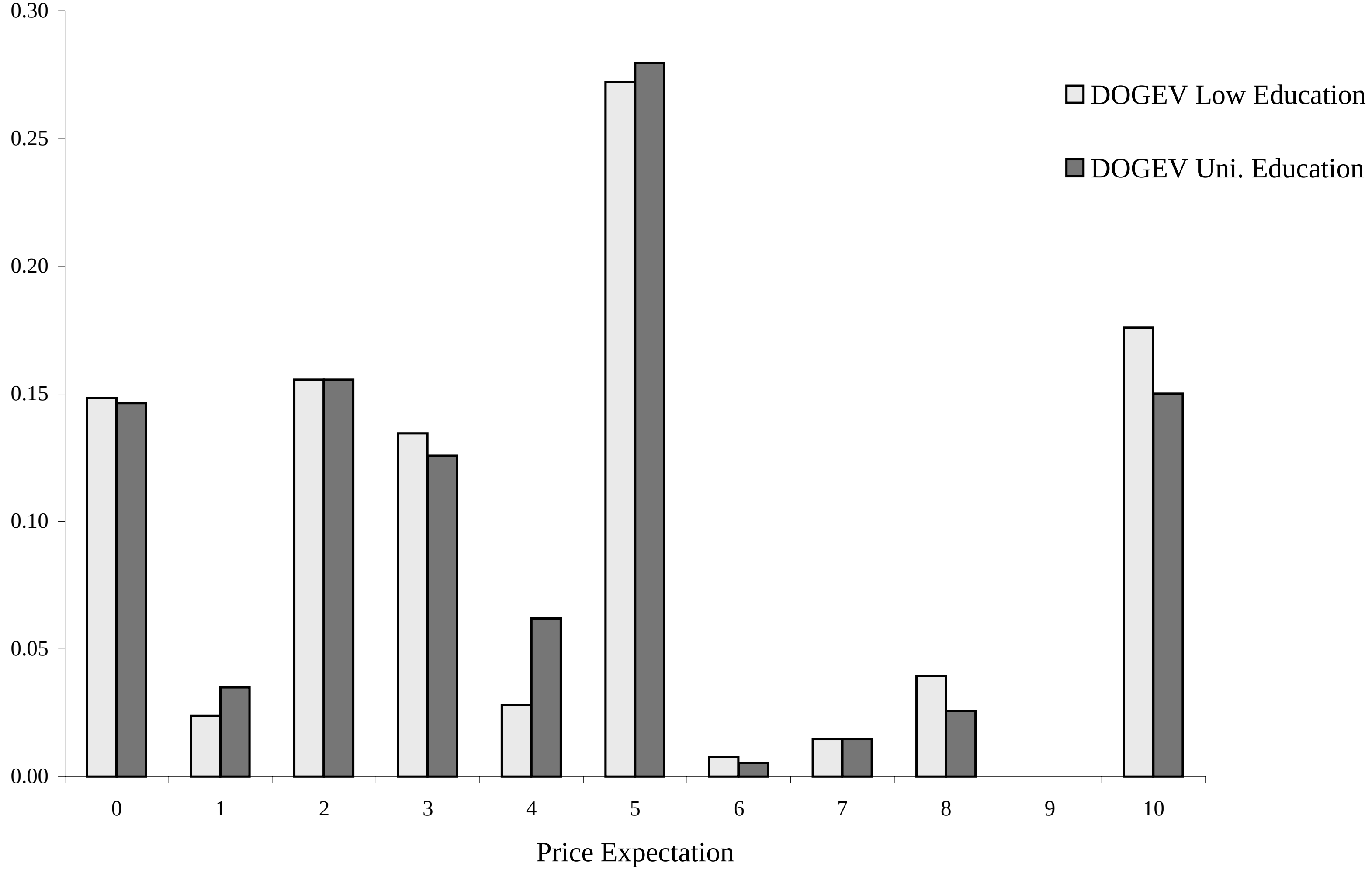

Figure 5: DOGEV Fitted Probabilities - Gender