Journal of Choice Modelling, 1(1), pp. 26-39
www.jocm.org.uk

Specification testing of discrete choice models: a note on the use of a nonparametric test

Mogens Fosgerau*

Institute for Transport, Technical University of Denmark, Knuth-Winterfeldts Allé, Bygning 116 Vest, 2800 Kgs. Lyngby, Denmark

Received 21 August 2007, revised version received 4 December 2007, accepted 7 January 2008

Abstract

Model misspecification is a serious issue since misspecification generally renders statistical inference invalid. However, specification testing of discrete choice models is rarely applied. This paper describes a nonparametric test procedure which uses a combination of smoothed residual plots and a test statistic able to detect general misspecification. Nonparametric methods require large datasets when the number of independent variables is more than a few. A way to circumvent this problem is indicated, increasing the usefulness of the approach also with limited datasets.

Keywords: discrete choice, specification test, nonparametric, functional form

1 Introduction

It is standard practice in regression models to perform model control using the residuals of the estimated model. Residuals are plotted to verify whether they are in fact white noise unrelated to the independent variables. Residuals are less easily defined in discrete choice models and similar model control is rarely performed for such models.

In fact, a variety of specification tests are available in the literature for some discrete choice models, but are not widely used. Lechner (1991) presents some specification tests for the binary logit model. Gourieroux et al. (1987a) and Gourieroux et al. (1987b) present tests based on generalised residuals for a range of models including the multinomial logit model. McFadden (1987) presents regression-based specification tests for the multinomial logit model similar in nature to the test presented here. The seminal paper by McFadden and Train

*T: +45 4525 6521, F: +45 4593 6533, mf@transport.dtu.dk

Licensed under a Creative Commons Attribution-Non-Commercial 2.0 UK: England & Wales License
http://creativecommons.org/licenses/by-nc/2.0/uk/

ISSN 1755-5345

(2000) provides specification tests of MNL and mixed logit against alternatives with more mixing. Finally, software exists that allows the comparison of observed choices to predictions when data are grouped on a categorical variable. This approach may be viewed as a kind of residual test.[1]

The point of this paper is to describe how the nonparametric test of functional form in Zheng (1996) may be applied to discrete choice models of general form. This means that the test applies even to complicated models such as the mixed generalised extreme value model. The Zheng test is based on nonparametric kernel regression of the parametric model residuals against the independent variables. With residuals defined as the difference between choice 0-1 indicators and predicted probabilities, the Zheng test applied to a discrete choice model is based on the comparison of predicted and observed choices. Pagan and Ullah (1999) review a range of nonparametric tests of functional form that are more or less similar to the Zheng test. The procedure presented in this paper of applying a test to a function of the independent variables is not restricted to the Zheng test but may be applied with other tests as well. On the use of nonparametrics in a discrete choice context and in transport applications see, e.g. Fosgerau (2006), Fosgerau (2007) and Fosgerau and Bierlaire (2007). Pagan and Ullah (1999), Yatchew (2003) and Härdle (1990) give general introductions to nonparametric and semiparametric methods.[2]

A general problem in using nonparametric techniques is that the demands on data increase exponentially in the dimension of the space of independent variables, this is the so-called curse of dimensionality. This issue is addressed in this paper by showing how such tests may be applied to a subset of variables or more generally to functions of variables such as the index representing the indirect utility of a choice alternative. It is thus possible to apply nonparametric tests to a model, while regressing only on a low number of variables or even just one. This reduces the demand on data and is of practical importance when the model to be tested has several independent variables.

Furthermore, the test may be applied to variables that are not included in the model. Thus the test may serve as a test of omitted variables.

The paper is organised as follows. Section 2 presents an exposition of the Zheng test. Section 3 shows how such tests may be applied to low-dimensional functions of the independent variables. Section 4 presents an example of application of the test to a discrete choice model while section 5 concludes.

2 The Zheng test

Consider a model $E(y|x)$, predicting the expectation of a variable $y \in \mathbb{R}$ conditional on an m-dimensional vector of observables $x \in \mathbb{R}^m$, which are also taken

[1] E.g. Alogit produces so-called apply tables.

[2] Tests based on regression of residuals against independent variables are also discussed in Ellison and Ellison (2000).

as random having density $p(x)$. The dependent variable y may be continuous, but it may also be binary or a binary dummy indicator of an alternative in a multinomial model. In such cases $E(y|x) = P(y = 1|x)$. In models with many alternatives it may also be useful to let y be a dummy indicator for the choice of a group of alternatives. This situation is also accommodated. In situations with unlabelled alternatives it could be useful to reorder the alternatives according to some characteristic such that y indicates, e.g. choice of the fastest alternative.

The expectation of y conditional on x is a function of x. We let $g(x)$ denote the true but unknown conditional expectation. The researcher specifies a parametric model $f(x : \theta)$ where $\theta \in \Theta$. We wish to test whether there is a parameter $\theta_0 \in \Theta$ such that $f(x : \theta_0) = g(x)$. Formally, we formulate our null hypothesis as

$$H_0 : P\left[f(x : \theta_0) = g(x)\right] = 1. \tag{1}$$

The null hypothesis thus states that $f(\cdot : \theta_0)$ coincides with g for almost all x. The corresponding alternative hypothesis is that, for all $\theta \in \Theta$, f differs from g on a set of probability greater than zero, or formally,

$$H_1 : \forall\theta \in \Theta \; P\left[f(x : \theta) = g(x)\right] < 1. \tag{2}$$

The alternative hypothesis is the negation of the null hypothesis. It thus encompasses all possible departures from the null.

Zheng (1996) defines a statistic T_n, where n is sample size, and shows that, under the null hypothesis, T_n converges in distribution to a standard normal as sample size increases, whereas it converges in probability to infinity under the alternative. Therefore the test is consistent against all departures from the parametric model f.

The idea of the test is the following. Define residuals by $\varepsilon = y - f(x : \theta_0)$. Then under the null hypothesis, $E(\varepsilon|x) = 0$ almost surely and hence also $E[\varepsilon E(\varepsilon|x)p(x)] = 0$. However, under the alternative hypothesis,

$$\begin{aligned} E[\varepsilon E(\varepsilon|x)p(x)] &= E[E(\varepsilon|x)^2 p(x)] \\ &= E[(g(x) - f(x : \theta_0))^2 p(x)] \\ &> 0. \end{aligned} \tag{3}$$

where the first equality follows from the law of iterated expectations.

Consider now an i.i.d. sample (y_i, x_i). The test statistic is formed from a sample analogue of $E[\varepsilon E(\varepsilon|x)p(x)] = 0$, constructed using kernel regression and kernel density estimation (see for example Pagan and Ullah 1999). In order to apply these methods we need a non-negative, bounded, continuous and symmetric kernel K with $\int K(u)du = 1$ and a bandwidth h depending on the sample size n. The application in section 4 uses a standard normal density for K. Then we construct first an estimate of the density of $x \in \mathbb{R}^m$ as

$$\hat{p}(x_i) = \frac{1}{n-1} \sum_{j \le n, j \ne i} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \tag{4}$$

and next an estimate of the expected residual as

$$\hat{E}(\varepsilon_i|x_i) = \frac{\frac{1}{n-1}\sum_{j\leq n, j\neq i}\frac{1}{h^m}K\left(\frac{x_i-x_j}{h}\right)\varepsilon_j}{\hat{p}(x_i)}. \tag{5}$$

Letting $\hat{\theta}$ be consistently estimated, define $e_i = y_i - f(x_i : \hat{\theta})$ and define next a sample analogue of $E[\varepsilon E(\varepsilon|x)p(x)]$ by

$$\frac{1}{n(n-1)}\sum_{i\leq n}\sum_{j\leq n, j\neq i}\frac{1}{h^m}K\left(\frac{x_i-x_j}{h}\right)e_i e_j \tag{6}$$

Zheng then defines a standardised version of the test statistic by

$$T_n = \frac{\sum_{i\leq n}\sum_{j\leq n, j\neq i}K\left(\frac{x_i-x_j}{h}\right)e_i e_j}{\left[\sum_{i\leq n}\sum_{j\leq n, j\neq i}2K^2\left(\frac{x_i-x_j}{h}\right)e_i^2 e_j^2\right]^{\frac{1}{2}}} \tag{7}$$

and shows that under some regularity conditions, if $h \to 0$ and $nh^m \to \infty$, then under the the null hypothesis, $T_n \to_d N(0,1)$, while under the alternative hypothesis,

$$\frac{T_n}{nh^{m/2}} \to_p c > 0. \tag{8}$$

Since T_n is a one-sided test statistic, we reject at level α if $T_n > Z_\alpha$, where Z_α is the upper α-percentile of a standard normal variable. For example, we reject H_0 at the 5 percent level if $T_n > 1.645$ (Li and Racine, 2007).

2.1 Some comments

Although the Zheng statistic looks complicated, there is a straight-forward intuition behind. The denominator in the expression for T_n takes care of the standardisation so we concentrate on the numerator. This is just a weighted sum of products of residuals.

The statistic is built up as a sum of contributions from each observation i. The weighting ensures that only observations near the i'th receive non-negligible weight. The weighted product $e_i e_j$ is positive if residuals e_j near e_i have the same sign as e_i. When the true and the estimated models are smooth functions of x, this will happen if

$$0 \neq g(x_i) - f(x_i : \hat{\theta}) = E(y_i - f(x_i : \hat{\theta})|x_i) = E(e|x_i). \tag{9}$$

The weighting ensures that local discrepancies between g and f are detected even when positive and negative discrepancies would net out globally.

The estimator in eq. (5) is a nonparametric kernel regression of the residuals ε against x. It is useful to use this regression to produce smoothed plots of

the residuals against the independent variables. Confidence bands around the regression can be computed using that

$$(nh)^{1/2}[\hat{E}(\varepsilon_i|x_i) - E(\varepsilon_i|x_i)] \sim N\left(0, \sigma^2(x_i)p^{-1}(x_i)\int K^2(u)du\right), \tag{10}$$

where σ^2 is the variance of ε conditional on x.[3] For the purpose of testing the specification in the parametric model f we may estimate this variance using the null hypothesis that $E(y|x) = f(x|\theta)$ a.e. In the case of a uni-dimensional x, we may estimate $\sigma(x)$ by

$$\hat{\sigma}^2(x) = \frac{\hat{f}(x)(1-\hat{f}(x))}{\hat{p}(x)nh}\int K^2(u)du, \tag{11}$$

where $\hat{f}$ is a kernel regression of the parametric model probabilities against x and $\hat{p}$ is the estimated density of x. For a normal density kernel, $\int K^2(u)du = \frac{1}{2\sqrt{\pi}}$.

The idea that the bandwidth h should depend on sample size may be unfamiliar. If the bandwidth is too large then averaging over too large neighbourhoods will smooth away relevant differences and hence introduce bias. If the bandwidth is too small then estimates will be noisy, they will be too much influenced by random fluctuations in data and hence variance will be high. Choosing the bandwidth appropriately as a function of sample size balances bias and variance and ensures that the bandwidth tends to zero at an appropriately slow rate, such that the asymptotical results obtain.

These considerations suggest that using a large bandwidth will tend to reduce the Zheng statistic as positive differences between f and g in some regions of the data will then be averaged with negative differences in other regions.

For a given sample size n, the bandwidth may be chosen by a rule such as, e.g. $h = n^{-\frac{1}{2m}}$. This choice is easy and agrees with the conditions for the Zheng test. It is also possible to select a bandwidth using cross-validation in the regression of the residuals against x. This is however time consuming and the gain from doing it is not clear.[4] In applied research, it may be sufficient or even preferable to select a bandwidth by just inspecting the resulting regression and the confidence bands visually, this is so-called eye-balling, suggested by Pagan and Ullah when the dimension of x is low enough to make this feasible.

3 Reducing dimensionality

A concern with the application of tests like the Zheng test, just as with any nonparametric technique, is the curse of dimensionality. The size of the dataset

[3] See Pagan and Ullah (1999).

[4] See the discussion in Li and Racine (2007). Generally all that is required for the test to be consistent is that bandwidths tend to zero as sample size tends to infinity but slowly enough that n multiplied by the product of bandwidths tends to infinity. This is, however, not very helpful in finite samples.

necessary to achieve a given degree of precision increases exponentially in the dimension of x. Rather than use the test directly it is therefore useful instead to consider the test defined over a function of the data.

Let $t(\cdot) : \mathbb{R}^m \to \mathbb{R}^{m_t}$ be a measurable function such that $t(x)$ has a density. Note that $m_t \leq m$, such that t is not in general injective. The inverse function of t is then a set-valued function defined by $t^{-1}(t') = \{x : t(x) = t'\}$. The idea is now to work with the m_t-dimensional stochastic variable t rather than with the m-dimensional stochastic variable x. So we apply the Zheng test to $E(y|t)$ rather than to $E(y|x)$. When $m_t < m$ there is a reduction of dimensionality and the test will be easier to apply. Note that

$$E(y|t_0) = E(y|x \in t^{-1}(t_0)) = E(E(y|x)|x \in t^{-1}(t_0)) \tag{12}$$

such that $g(t) = E(y|t)$ and $f(t : \theta) = E(f(x : \theta)|t)$ are the corresponding true and modelled conditional means, now defined over t rather than over x.

It is straight-forward that our null hypothesis H_0 implies the version reformulated in terms of functions t. If the model is true, then the model is also true as a function of any function t.

$$H_0 \Rightarrow H_0^* : \exists \theta_0 \, \forall t \; P\left[f(t : \theta_0) = g(t)\right] = 1. \tag{13}$$

It is also clear that the converse hypothesis H_1 implies the converse of H_0^*. That is, if the model is false, then there exists a function t such that the model is false as a function of t.

$$H_1 \Rightarrow H_1^* : \forall \theta \in \Theta \, \exists t \; P\left[f(t : \theta) = g(t)\right] < 1. \tag{14}$$

To verify the latter, simply choose a function t as $t(x) = g(x) - f(x : \theta)$. The reverse implications also hold: H_0^* implies the converse of H_1^*, which implies the converse of H_1, which in turn implies H_0. So H_0 is equivalent to H_0^* and H_1 is equivalent to H_1^*.

3.1 Some comments

It is now possible to proceed by selecting various functions t of the data and to conduct the Zheng test (or another similar test) on the model conditional on these t's. We cannot be sure to reject a false model for any given t, as the discrepancy between the parametric model and the true model may not lie in the direction of t. But if we reject the model for any t, then we are under the alternative hypothesis H_1^* and hence also under H_1, which means that the model has been rejected. Otherwise the model is accepted. The test based on functions t will be less powerful since we may not be able to choose appropriate t's. On the other hand we gain the advantage that the dimension of t can be chosen to be low, which greatly facilitates the use of nonparametric kernel methods.

Note that it is not necessary that every variable in x actually appears in the model f. If a variable x_o is omitted from the parametric model we may define $t(x) = x_0$ to form a nonparametric omitted variable test.

In a discrete choice model where the systematic utility of an alternative is defined by βx with β fixed, we may test the model by defining $t(x) = \beta x$ and take y as the binary indicator for choice of the same or another alternative.

When the dimension of t is 1 or 2 it is feasible to select the bandwidth visually by plotting the nonparametric regression of the residuals against $t(x)$ and varying the bandwidth until an appropriate number of features arise. It may be informative to inspect the data visually in this way and the Zheng test may then be used to check if detected deviations of the expected residual from zero are in fact significant.

It is hard to very specific about how this should be done. But usually one has some sense of how complicated a potential relationship might be. A low bandwidth tends to lead to an estimated relationship with many peaks. If there are more peaks than one thinks is likely in the true relationship, then it is appropriate to increase the bandwidth. On the other hand, choosing a too large bandwidth will smooth away many differences. This may introduce bias so it might be better to use a somewhat smaller bandwidth.

4 Examples

4.1 Mode choice

This section provides an illustrative application of the test to a mode choice model. The model is not intended as a serious model. It merely serves the purpose of showing how the Zheng test may be used to detect significant differences between model and data. I use a mode choice data set from Sweden comprising 1799 observations of trips to Stockholm using either air, bus, car or train. For each observation the data give travel time and cost for each alternative as well as some background variables.

I estimated an MNL with alternative specific constants, mode-specific parameters for travel time and a common cost parameter. Write this model with utilities $U_i = V_i + \varepsilon_i$, where ε_i are iid. extreme value and $V_i = \beta x_i$. Then I computed the predicted mode choice probabilities for the observations in the sample. Residuals were computed for each mode by subtracting the predicted probability from a dummy indicator for the observed choice. All the tests are carried out using uni-dimensional kernel regressions and density estimates were carried out using a normal density kernel and a bandwidth of $n^{-1/2}$, where n is sample size and data ($t(x)$) are scaled to the unit interval. The choice of kernel is generally not important. Judging from the graphs in the following, the bandwidth seems appropriate. For computing confidence intervals around the kernel regression I used equation (10).[5] I trimmed away 1 percent of the sample, excluding the up-

[5]The MNL was estimated in Biogeme (Bierlaire, 2003, 2005), which was also used to compute the predicted choice probabilities. The Zheng test and the nonparametric regressions were performed in Ox (Doornik, 2001). The Ox code for the test is given in the appendix.

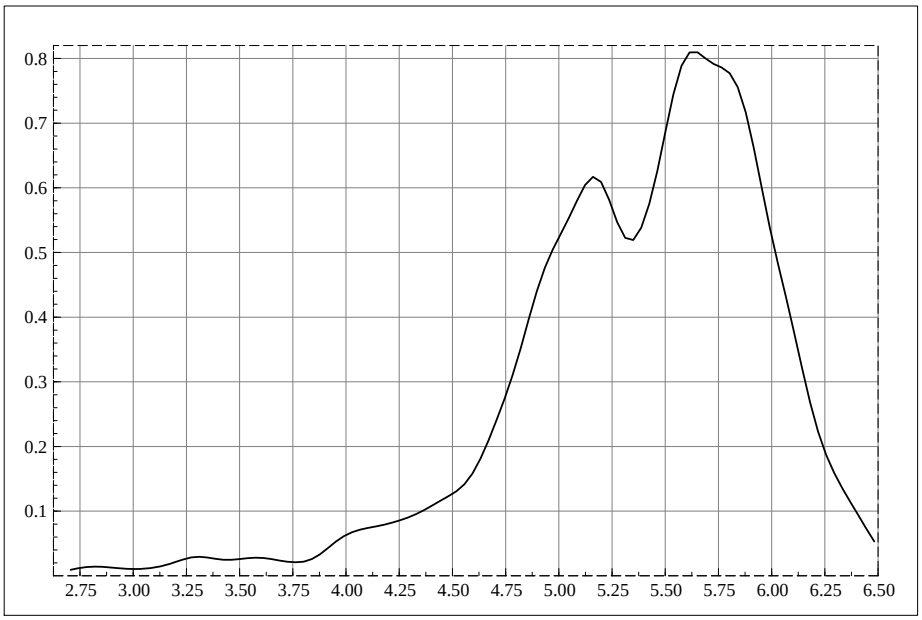

Figure 1: Density of log of real income

per and lower 0.5 percent after sorting on the independent variable, in order to exclude regions with very thin data and hence uncertain estimates.

4.1.1 Omitted variable test

An income variable is available in the data but has not been included in the model. Suppose we want to test whether income could improve the prediction of choices. In this simple model one could of course just try estimating some specifications including an income variable. For more complicated models it could however be prohibitively time-consuming to estimate many different versions of the same model. In such cases an omitted variable test could be useful.

Figure 1 shows first a kernel density estimate of the log of real income. The Zheng statistics for the four alternatives are 3.3 (Air), 0.7 (Bus), 1.1 (Car) and 3.7 (Train), which means that the residuals depend systematically on income for the Air and Train alternatives. Figures 2 and 3 show the corresponding plots for the air and bus modes. It is clear that the residuals of choosing air increases with income. Thus there is a dependency of choices on income that is not captured by the model. A similarly clear relationship is not visible for the bus residuals.

The dependency on income is, of course, unsurprising. The point here is that the test is well able to detect it. Income may be included in the model in various ways and the test does not indicate how it should be done. The smoothed residual plots suggest a specification that is linear in log income. It is

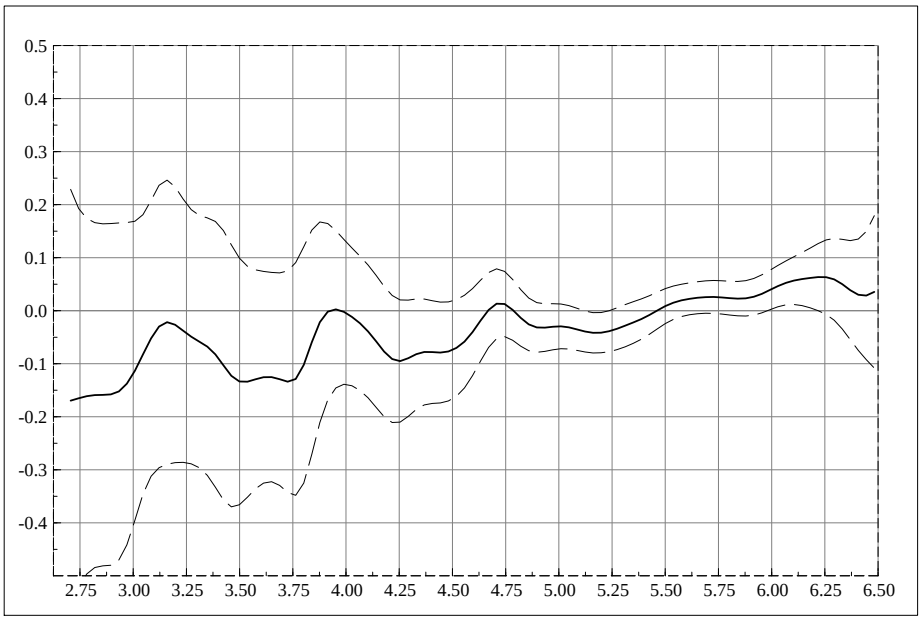

Figure 2: Residuals of air alternative on real income

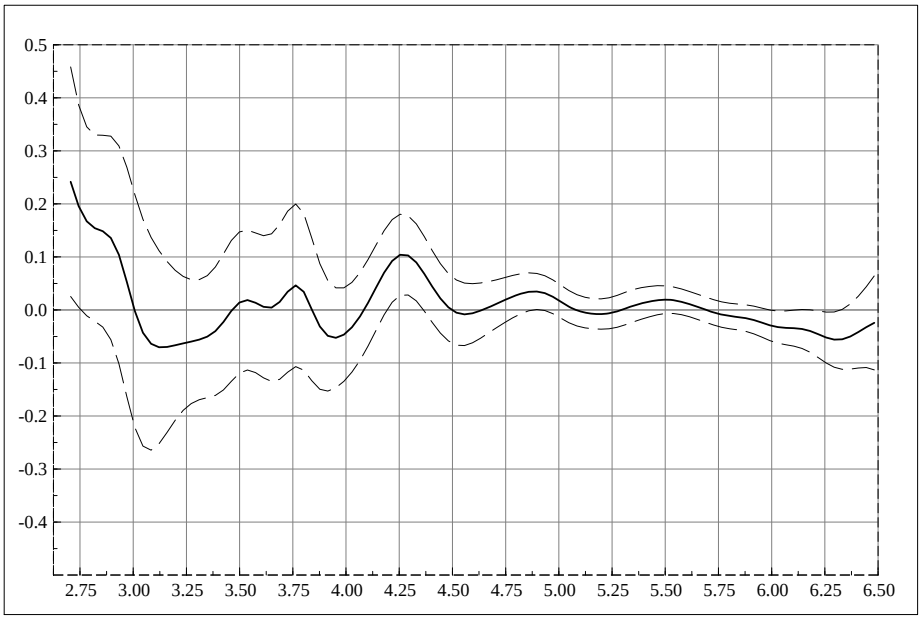

Figure 3: Residuals of bus alternative on real income

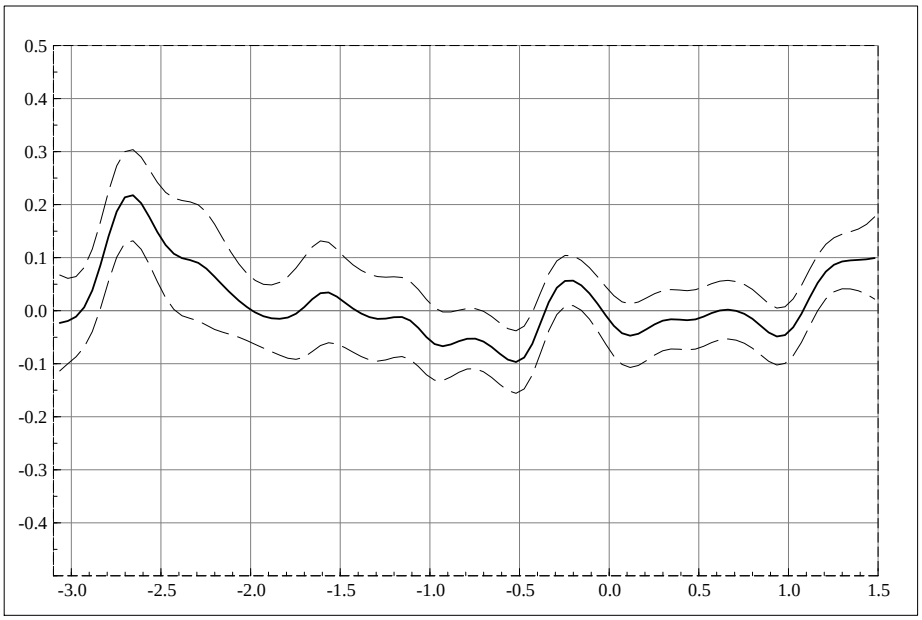

Figure 4: Residuals of train alternative on systematic utility for bus

not possible to conclude from the test which alternatives income should enter as an income variable in one alternative would affect the predicted probabilities of all alternatives. I tested a model formulation with separate income parameters for air, car and train and found those for air and car to be confidently positive, while that for train to be insignificant.

4.1.2 Misspecification test

It depends on the application whether exclusion of a significant omitted variable can be considered a misspecification. But a significant value of the Zheng statistic for variables that are included in the model must be considered evidence of misspecification. Table 1 shows the Zheng statistic for residuals of the four modes on the systematic utilities for the four modes. The test does not find significant misspecification over the systematic utility of the air mode, while there are systematic differences along the directions of the systematic utilities of the remaining modes. Figure 4 shows residual plots for the Train residuals on the systematic utility for Bus.

Judging from the estimated parameters and the loglikelihood it was impossible to detect whether the model is misspecified. But the test clearly shows that the model is indeed misspecified. The test is general against misspecification and so does not indicate the cause of misspecification. In a serious application it would therefore be pertinent to seek to elaborate or otherwise change the model.

Table 1: Zheng statistics for tests on systematic utilities V

	Air	Bus	Car	Train
V(Air)	0.2	0.8	-0.2	0.0
V(Bus)	3.2	3.7	1.9	6.0
V(Car)	5.5	9.7	2.7	2.4
V(Train)	5.0	3.4	1.12	2.4

4.2 Demand for alternative vehicles

The example in this section is from McFadden and Train (2000), who illustrate specification testing of the mixed logit model using a dataset on the demand for alternative-fuelled vehicles. They take an existing model and apply a test to detect that further mixing of some parameters is required.

Their dataset is available on the website of the Journal of Applied Econometrics. It has 4654 respondents who were asked to choose among six alternatives characterised by price, range, acceleration, top speed etc. I reproduced the estimates of McFadden and Train's final model[6] and increased the number of draws used in the simulation to 500 Haltons. The model has a linear specification with 22 independent variables and 22 parameters for each alternative, six of which are random with a normal distribution.

The only variable in the model with sufficiently many different values to be called continuous is the price variable. I therefore first conducted the Zheng test against the predicted probability for each alternative. The Zheng test statistics were all very large, ranging from 10.1 to 81.6 for the residuals on the first alternative. Moreover, the residual plots revealed that residuals for each alternative tended to all have the same sign. I therefore included alternative specific constants in the model. They turned out to be extremely significant. The likelihood improved from -7360.57 to -6977.27.[7]

Testing the model again still revealed evidence of some misspecification. The Zheng statistics for residuals on the probability for alternative 1, e.g. ranged up to 3.1, while the Zheng statistics for residuals on the price of the first alternative did not indicate any misspecification. Repeating the test with an increased bandwidth did, however, indicate misspecification. Inspection of the residual plots indicated many cases where residuals seemed to follow a monotonous relationship with the independent variable. I therefore extended the model by allowing for full correlation between the six random parameters. This yielded a significant improvement of the likelihood to -6969.95. From this point I did not find further evidence of misspecification.

[6] Except some cases where parameters were off by factors of ten.

[7] These constants indicated that the first and third alternatives were preferred by respondents. This could be an effect of the presentation of the choice situations or an effect of the design. I have not pursued this issue.

5 Concluding remarks

The issue of potential model misspecification should be taken very seriously, since statistical inference from a misspecified model is generally not valid. This paper has pointed out how an existing nonparametric test of functional form, the Zheng test, may be applied to discrete choice models in order to detect general misspecification.

The curse of dimensionality is a general problem with nonparametric methods and limits their usefulness. The paper points out how dimensionality may be reduced by applying the test not to the raw data but to functions of the data. This opportunity is not unique to the Zheng test but may be used with other similar test.

Reducing the dimensionality of the data by applying the test to a function of the data makes it a lot easier to apply the test and data requirements are reduced. The associated cost is that the result is conditional on the chosen function, since any given function may not be able to reveal a given misspecification. So while significant rejection of the model using a given function of the data is conclusive evidence against the model, it is possible that some misspecification will remain undetected. It is therefore advisable to perform the test using several functions of the data.

Application of the test is not difficult but requires some programming. In the appendix I have included the Ox function that I used to compute the Zheng statistic. The test has also been included in Biogeme along with facilities to produce graphs, making the test very easy to use within that environment.

Acknowledgments

Mogens Fosgerau has received financial support from the Danish Social Science Research Council.

A Ox code for Zheng test

I have here included a function in Ox that takes a column of unidimensional x, a column of unidimensional residuals and a bandwidth to compute the Zheng statistic. The bandwidth relates to data in the unit interval, the function scales the bandwidth to the actual range of the data.

```
DoZheng(const x, const res, const bw)
{
    decl Tnumerator, Tdenominator, Ttemp, T, i, bandwidth, density;
    bandwidth = bw * (maxc(x) - minc(x));

    Tnumerator = 0;
```

```
    Tdenominator = 0;

    for (i = 0; i < rows(x); i++)
    {
        density = densn((x[i] - x')./bandwidth);
        Ttemp = ( density .* res') * res[i] ;
        Ttemp[][i] = 0;
        Tnumerator = Tnumerator + sumr(Ttemp);
        Ttemp = 2 * ((density .^ 2) .* (res').^2) * (res[i].^2) ;
        Ttemp[][i] = 0;
        Tdenominator = Tdenominator + sumr(Ttemp);
    }
    Tdenominator = sqrt(Tdenominator);
    T = Tnumerator/Tdenominator;
//  "T test value ~ N(0,1) is : ", T;
    return T;
}
```

References

Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. Proceedings of the 3rd Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland.

Bierlaire, M., 2005. An introduction to Biogeme. biogeme.epfl.ch.

Doornik, J. A., 2001. Ox: An Object-Oriented Matrix Language. Timberlake Consultants Press, London.

Ellison, G., Ellison, S. F., 2000. A simple framework for nonparametric specification testing. Journal of Econometrics 96 (1), 1–23.

Fosgerau, M., 2006. Investigating the distribution of the value of travel time savings. Transportation Research Part B: Methodological 40 (8), 688–707.

Fosgerau, M., 2007. Using nonparametrics to specify a model to measure the value of travel time. Transportation Research Part A: Policy and Practice 41 (9), 842–856.

Fosgerau, M., Bierlaire, M., 2007. A practical test for the choice of mixing distribution in discrete choice models. Transportation Research Part B: Methodological 41 (7), 784–794.

Gourieroux, C., Monfort, A., Renault, E., Trognon, A., 1987a. Generalised residuals. Journal of Econometrics 34 (1-2), 5–32.

Gourieroux, C., Monfort, A., Renault, E., Trognon, A., 1987b. Simulated residuals. Journal of Econometrics 34 (1-2), 201–252.

Härdle, W., 1990. Applied Nonparametric Regression. Vol. 19 of Econometric Society Monograph Series. Cambridge University Press.

Lechner, M., 1991. Testing logit models in practice. Empirical Economics 16, 177–198.

Li, Q., Racine, J. S., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton and Oxford.

McFadden, D., 1987. Regression-based specification tests for the multinomial logit model. Journal of Econometrics 34 (1-2), 63–82.

McFadden, D., Train, K., 2000. Mixed mnl models for discrete response. Journal of Applied Econometrics 15, 447–470.

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press.

Yatchew, A., 2003. Semiparametric Regression for the Applied Econometrician. Themes in modern econometrics. Cambridge University Press.

Zheng, J. X., 1996. A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics 75 (2), 263–289.