The Journal of Choice Modelling 21 (2016) 60–65

Contents lists available at ScienceDirect

The Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Discrete choice models' ρ^2: A reintroduction to an old friend

Patricia L. Mokhtarian

School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA 30332, USA

ARTICLE INFO

Article history:
Received 10 August 2015
Received in revised form
4 January 2016
Accepted 1 February 2016
Available online 10 August 2016

Keywords:
McFadden's R^2
Pseudo-R^2
Goodness-of-fit
Unequal choice sets
Segmentation

ABSTRACT

We first review the intuition behind ρ^2, and its conceptual interpretation. We then comment on the choice of benchmark (typically either the equally-likely, EL, or the market-share, MS, model), together with discussion of what is considered a "good" value. After a brief mention of the adjusted ρ^2 and of statistical distributions associated with ρ^2, we close with a description of its computation under three special circumstances: repeated observations, unequal choice sets, and deterministically-segmented models.

1. Introduction

The quantity (misleadingly) known as "McFadden's R^2" (because proposed by McFadden, 1973), but which we will refer to as ρ^2 (after Ben-Akiva and Lerman, 1985; Train, 2009), is probably the most commonly-used measure of goodness of fit for discrete choice models (Veall and Zimmermann, 1996). Yet it is rarely discussed at any length in textbook treatments of the subject. This brief note assembles in one place a number of facts and perspectives about ρ^2, drawn from diverse published and unpublished sources. The goal is to provide applied researchers and model users a more complete understanding of the concepts and practice associated with this important measure.

2. Conceptual interpretation

The log-likelihood $\mathcal{L}(\boldsymbol{\beta})$ for a discrete choice model (DCM) with coefficients $\boldsymbol{\beta}$ is bounded on both sides. The effective lower bound is $\mathcal{L}_{\text{EL}} = \mathcal{L}(\mathbf{0})$, the case of no information, in which all coefficients are zero and all alternatives are considered to be equally likely (EL). It is possible to obtain a lower $\mathcal{L}$, but as Train (2009) points out, coefficients in any model with $\mathcal{L}(\boldsymbol{\beta}) < \mathcal{L}_{\text{EL}}$ cannot be maximum likelihood estimators, since by construction there is clearly a set of coefficients (the zero vector yielding the EL model) with a higher $\mathcal{L}$. The theoretical upper bound is $\mathcal{L}_{\max} = 0$, the case of perfect information, in which the chosen alternative is always predicted with probability 1 (it is only a theoretical bound, since achieving a perfectly-predicting model would generally require as many parameters as observations, which would not be applicable beyond the estimation sample).

Since the log-likelihood for any DCM will fall between these two bounds, it makes sense to define a goodness-of-fit measure in terms of how far along that continuous interval the $\mathcal{L}$ for a given model is. That is, what proportion of the

E-mail address: patmokh@gatech.edu

http://dx.doi.org/10.1016/j.jocm.2016.02.001
1755-5345/

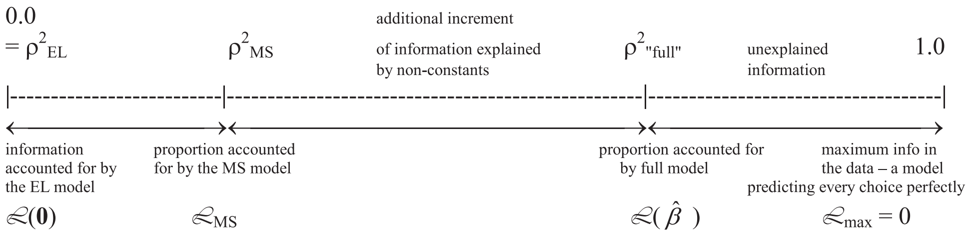

Fig. 1. Incremental increases in proportion of information explained by a series of nested models estimated on the same data (Source: adapted from Hauser (1978)).

distance from worst case to best case is "covered" or "explained" by that model? The higher that proportion (i.e. the closer $\mathcal{L}$ is to the best case of 0), the better the model (in terms of fit – conceptual considerations aside). That proportion is (see Fig. 1):

$$\left[\mathcal{L}(\hat{\boldsymbol{\beta}}) - \mathcal{L}(\mathbf{0})\right] / \left[\mathcal{L}_{\max} - \mathcal{L}(\mathbf{0})\right] = \left[\mathcal{L}(\mathbf{0}) - \mathcal{L}(\hat{\boldsymbol{\beta}})\right] / \mathcal{L}(\mathbf{0}) = 1 - \left[\mathcal{L}(\hat{\boldsymbol{\beta}}) / \mathcal{L}(\mathbf{0})\right]. \tag{1}$$

This measure is sometimes referred to in the older literature as the *likelihood ratio index,* or *LRI*, not to be confused with the χ^2 *likelihood ratio test* of a constrained versus less-constrained model. It is one of a number of quantities known as "pseudo-R^2" measures (so in reporting model results, it is important to specify *which* pseudo-R^2 measure is being used). As with other pseudo-R^2 measures, ρ^2 ranges between 0 (when $\mathcal{L}(\hat{\boldsymbol{\beta}}) = \mathcal{L}(\mathbf{0})$) and 1 (when $\mathcal{L}(\hat{\boldsymbol{\beta}}) = 0$, the maximum possible).

Because it shares the same bounds, as well as the conceptual interpretation as a measure of "proportion of something explained by the model" and "improvement from a null model to a fitted model" (http://statistics.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm, accessed August 2, 2015), ρ^2 is "analogous to the multiple-correlation coefficient [R^2] in the linear statistical model" (McFadden, 1973, p. 121). However, their interpretations, while conceptually similar, are technically different. Recall that

$$R^2 = \text{Regression Sum of Squares/Total Sum of Squares} = \frac{\sum_n (\hat{Y}_n - \bar{Y})^2}{\sum_n (Y_n - \bar{Y})^2} \tag{2}$$

(where Y_n is the observed value of the dependent variable Y for the nth case, $\hat{Y}_n$ is the corresponding value predicted by the estimated model, and $\bar{Y}$ is the sample mean of Y), which is interpreted as the *proportion of total variance in the dependent variable Y that is explained by the model.* ρ^2, on the other hand, is clearly a proportion, but the denominator is not a measure of total variance, it is an initial log-likelihood. Hence, it is interpreted rather literally by Ben-Akiva and Lerman (1985), p. 91), simply as "the fraction of an initial log-likelihood value explained by the [full] model".

However, ρ^2 does have an elegant interpretation as the proportionate reduction in the total entropy or uncertainty of the data as defined by a benchmark "null" model (such as the equally-likely or market-share model), that is achieved by the full model. Hauser (1978) shows that this total uncertainty (as defined by a given benchmark) can also be viewed as the total amount of information (starting from the same baseline) that is available to be explained by a perfect model. Therefore, ρ^2 is the empirical *proportion of information in the choice data (as defined by the benchmark null case) that is explained by the model* (Hauser, 1978; Cameron and Windmeijer, 1997). It is a measure of the predictive ability of the model, in that better models will tend to have higher predicted probabilities of the chosen alternatives (culminating in the theoretical perfect model, for which those predicted probabilities are identically equal to 1), which means greater information explained, or lower entropy or uncertainty (where the perfect model, explaining all information, has zero entropy).

Incremental improvements in $\mathcal{L}$ obtained by adding variables to a model can be graphically portrayed as increases in the proportion of information explained, as shown in Fig. 1 which can alternatively be viewed as illustrating deviance decomposition (Veall and Zimmermann, 1996). Cameron and Windmeijer (1997) propose a generalization of ρ^2 (with a constant-only model as base) applicable to other exponential-family models, such as Poisson and gamma. Their measure represents the "proportionate reduction in recoverable information" (p. 333) or "in uncertainty" (p. 329), due to the model. They comment (p. 333) that their measure "can be interpreted as being based on deviance residuals", which "have been found very useful for diagnostic checking", and thus that it is "related to the analysis of deviance the same way as R^2 in the standard linear model is related to the analysis of variance." Menard (2000, p. 24) also supports the interpretation of ρ^2 as the best conceptual counterpart to the regression R^2, in the sense of representing "a proportional reduction in error measure".

3. Benchmarking the ρ^2

The benchmark or denominator of Eq. (1) is the worst-case, equally likely (EL) model, but we could also have used some other reference model (generally a constrained, or "nested", version of the final model of interest) to obtain an initial log-likelihood. A popular alternative is to use the market share (MS) model as the benchmark, i.e. a model with only constant term(s), for which the predicted probability of choosing a given alternative is equal to the market share of that alternative for each case in the sample. A number of reasons for this popularity can be identified.

First, it can be argued that we are never so ignorant about a choice context that we must resort to assuming all choices are equally likely; we always know at least the market shares of each alternative (in our sample). Second, starting with the MS model is a more conservative practice, since the same final $\mathcal{L}$ will cover a smaller fraction of the distance from $\mathcal{L}_{\text{MS}}$ to 0 than of the distance from $\mathcal{L}_{\text{EL}}$ to 0. Third, whereas the EL-base ρ^2 for the MS model will be some positive number (unless the market shares are equal, in which case the MS model is the same as the EL model), there is no counterpart to an EL-base ρ^2 for regression, and the R^2 for the constant-only ("MS") regression model is 0. To see this, recall that for a regression model with just a constant term, the parameter estimate for the constant term is just $\bar{Y}$, $\hat{Y}_n$ is identically equal to $\bar{Y}$ for all n, and the Regression Sum of Squares in Eq. (2) is identically 0. Thus, in regression the worst-case benchmark *can only be* the MS model (Tardiff, 1976). Finally, for nested logit models there is some ambiguity about how the equally-likely model is defined. In Limdep, for example (Greene, 2009), the "zero coefficients" $\mathcal{L}$ represents an assumption of equal probabilities among alternatives within each nest, and equal probabilities for each nest, which is not the same as equal probabilities for each elemental alternative. Thus, in this case $\mathcal{L}_{\text{EL}}$ will differ depending on the nesting structure being modeled, which would make it problematic to use an EL-based ρ^2 to informally compare models. By contrast, the market shares are invariant to the nesting structure assumed, so an MS-based ρ^2 would allow an apples-to-apples comparison.

With the weight of these excellent arguments, plus the authority of Ortúzar and Willumsen (2011), who advocate the exclusive use of the MS base for ρ^2 (referring to it, on p. 282, as the *corrected* ρ^2), why does this author prefer to keep the EL model as base (while still comparing $\mathcal{L}(\hat{\boldsymbol{\beta}})$ to the MS log-likelihood $\mathcal{L}_{\text{MS}}$ as well, as will be elaborated below)? Let me explain.

In general, the EL-base ρ^2 for the MS model will be smallest (0) when the shares are equal, and increase with the disparity in shares. Thus, when shares are unbalanced, even the MS model will appear to have excellent fit when using the EL benchmark. For example, with two alternatives and shares equal to (0.1, 0.9), the EL-base ρ^2 for the MS model is 0.53 (from Eq. (1), with $\mathcal{L}_{\text{EL}} = -N \ \ln 2$ and $\mathcal{L}_{\text{MS}} = N_1 \ln(N_1/N) + N_2 \ln(N_2/N)$, where N_j = number of cases choosing the jth alternative, and $N_1+N_2=N$). When the MS-model ρ^2 is so high, it can be quite difficult for real variables (i.e. non-constants) to add much explanatory power. Imagine that we add some real variables to the MS model just described, and that they are statistically significant and conceptually meaningful, but the final model (EL-base) ρ^2 is, say, 0.54! Virtually no improvement at all, it appears, and *if the MS-base ρ^2 had been used*, the final model would have had a very disappointing ρ^2 of nearly 0! Should we be indifferent between the two models, since their final $\mathcal{L}$s are so similar? Should we even prefer the MS model, on the grounds of parsimony? Should we despair that the fortune we spent collecting data on real variables was thrown away? Almost certainly not: we would generally prefer the model with real variables, which offers *behavioral insight into why the market shares are what they are*. Giving such a model a ρ^2 of about 0 sets the bar too high and is an unfair indication of its true value.

We should realize that adding real variables to a MS model does more than just provide *incremental* explanatory power; rather, such variables are also "robbing" the constants of some (perhaps most or all) of *their* "explanatory" power. The real variables should "get credit" accordingly, which is the logic behind starting from the EL benchmark. Especially in cases where shares are unbalanced, I recommend reporting the $\mathcal{L}$ for an otherwise final model *without the constant term(s)* (as well as the standard $\mathcal{L}_{\text{EL}}$, $\mathcal{L}_{\text{MS}}$, and final $\mathcal{L}(\hat{\boldsymbol{\beta}})$), to reveal the full explanatory contribution of the real variables. This would never be the final model, since the constants are needed to ensure (for multinomial logit) that predicted shares replicate observed shares (Ben-Akiva and Lerman, 1985) and to absorb the effects of choice-based sampling (Manski and Lerman, 1977). But this proposal simply shifts the perspective from seeing only the incremental contribution of the real variables on top of the constants, to seeing the *incremental contribution of the constants (representing the average impact on utility of "everything unobserved")* ***after*** *accounting for the role of the observed and behaviorally-meaningful variables.*

By the same token, however, it is not appropriate to report *only* the $\mathcal{L}_{\text{EL}}$ and final $\mathcal{L}(\hat{\boldsymbol{\beta}})$, since that could obscure the relative importance of the constant terms and make the model look much better than the real variables alone would support. A market share benchmark is always important.

From the foregoing discussion, it is clear that a poor model can have a high EL-base ρ^2 (due to unbalanced market shares) and conversely a model with a lower ρ^2 may still be a big improvement over the MS model (when shares are nearly even). Can we say anything, then, about "how big a ρ^2 would be considered 'good' for a model?" Ben-Akiva and Lerman, 1985, p. 167) state, "There are no general guidelines for when a ρ^2 value is sufficiently high." Veall and Zimmermann (1996, p. 256) advise that evaluations of ρ^2 should "be informed by the modeling context, just as researchers expect OLS-R^2 to be larger with aggregate time series data in levels than with cross-sectional data."

Hensher et al. (2005) go a bit farther, presenting a figure (taken from Domencich and McFadden, 1975) that empirically relates ρ^2 to R^2. However, neither source specifies which base was used for ρ^2, EL or MS, and as we have just seen, it makes a difference. Domencich and McFadden acknowledge both possible bases, but do not mention which they used for the figure, although for a comparison to R^2 it would be natural to use the MS base, for reasons mentioned above. Furthermore, neither source provides the evidence giving rise to the figure. In general, Hensher et al. (2005, pp. 338–339) appear to prefer the MS base, which is presumably the base assumed in their comment that "in our experience, a pseudo-R^2 of 0.3 represents a decent model fit for a discrete choice model" (corresponding to an R^2 of about 0.6, according to the figure). Although the difficulty of achieving a fit of 0.3 will vary depending on the number of alternatives and their shares, a fit of that size would be probably considered respectable by most DCM users.

4. The adjusted ρ^2

Since ρ^2 can be made arbitrarily high with the addition of more variables, Ben-Akiva and Lerman (1985) suggest the use of an adjusted ρ^2 measure, which penalizes the model for lack of parsimony, analogous to the adjusted R^2 of regression. For the EL base we have

$$adj.\ \rho^2 = 1 - [\mathcal{L}(\hat{\beta}) - K]/\mathcal{L}_{EL}, \tag{3}$$

and for the MS base

$$adj.\ \rho^2 = 1 - [\mathcal{L}(\hat{\beta}) - K + J - 1]/\mathcal{L}_{MS}, \tag{4}$$

where K is the number of estimated parameters and J is the number of alternatives (so that $K-(J-1)$ is the number of parameters other than constant terms being estimated, and when the MS model is the base, the full model is penalized only for the *additional* parameters besides the constants). The motivation for this adjustment is that the numerator of the fraction in Eq. (3), known as the Akaike Information Criterion (AIC), is the expected value, taken over *all possible* samples, of the log-likelihood based on the $\hat{\beta}$s estimated on *this* sample (properly correcting $\mathcal{L}(\hat{\beta})$ for the fact that this particular set of $\hat{\boldsymbol{\beta}}$s will generally not produce as high a log-likelihood on some other sample as it did for this one, for which it maximized the log-likelihood). Choosing a model that maximizes the adj. ρ^2 is equivalent to maximizing the AIC, which chooses the $\hat{\boldsymbol{\beta}}$s (among a set of candidates) that are expected to perform the best across all samples (Ben-Akiva and Swait, 1986).

The adjusted ρ^2 can decrease with the addition of new variables; it is easy to show (for either formulation) that it will *increase* if and only if the estimation of M new parameters increases the final log-likelihood by more than M. For large samples, $\mathcal{L}(\hat{\boldsymbol{\beta}})$ and $\mathcal{L}_{\mathrm{EL}}$ or $\mathcal{L}_{\mathrm{MS}}$ will tend to be large relative to K, and thus adj. ρ^2 will tend to differ very little from ρ^2. The adjusted ρ^2 is also the basis for a statistical test comparing non-nested models, i.e. a pair of models for which neither is a constrained version of the other (see Ben-Akiva and Lerman, 1985, p. 172; Ben-Akiva and Swait, 1986).

5. Relationship of ρ^2 to statistical distributions

The ρ^2 measure itself has no known statistical distribution, but Domencich and McFadden, 1975, pp. 123–124) note that

$$\left(\frac{K}{K-\bar{K}}\right)\left(\frac{\rho^2}{1-\rho^2}\right) \sim_{asympt} F(K-\bar{K}, K), \tag{5}$$

where K is the number of parameters in $\hat{\boldsymbol{\beta}}$ and $\bar{K}$ is the number of parameters in the base model (0 if EL, number of alternatives less 1 if MS), and (analogously to the F-test for significance of regression) "this distribution can be used to test the hypothesis $\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$" (where the latter is the $\boldsymbol{\beta}$ vector for the base model, i.e. all 0 s for EL, and only the alternative specific constant coefficients for MS). However, ρ^2 is most often used for qualitative comparisons rather than for tests of statistical significance. The χ^2 likelihood ratio test is most often used for the latter purpose.

Following Windmeijer (1995, p. 105; also see Tardiff, 1976, p. 380), we observe that ρ^2 (using the EL model as a base) "has a one-to-one relation with the chi-squared statistic for testing the hypothesis that all coefficients … are zero." This follows directly from Eq. (1), which defines ρ^2_{ELbase}:

$$-2\left[\mathcal{L}(\mathbf{0})\right]\rho^2_{\mathrm{ELbase}} = -2\left[\mathcal{L}(\mathbf{0}) - \mathcal{L}(\hat{\boldsymbol{\beta}})\right] \sim_{\mathrm{asympt.}} \chi^2_{K\ \mathrm{df}} \text{ under } \mathrm{H}_0\text{: } \beta_k = 0,\ k = 1, \ldots, \mathrm{K}. \tag{6}$$

I use this expression as the basis for a series of homework exercises in my DCM class. Since, when all decision-makers have the same J alternatives in their choice set, $\mathcal{L}(\mathbf{0}) = -N \ln J$, specifying K and α allows us to find the critical χ^2 value, and thence the critical ρ^2 value, below which a model would not be statistically better than the EL model. More interestingly, let us temporarily adopt the MS-base ρ^2, and (for simplicity) take $J=2$. Then by the same algebra, we have

$$-2\left[\mathcal{L}_{\mathrm{MS}} - \mathcal{L}(\hat{\boldsymbol{\beta}})\right] = -2\,\mathcal{L}_{\mathrm{MS}}\,\rho^2_{\mathrm{MSbase}} \sim_{\mathrm{asympt.}} \chi^2_{K-1} \text{ under } \mathrm{H}_0\text{: } \beta_k = 0,\ k = 2, \ldots, \mathrm{K}, \tag{7}$$

where β_1 is the constant term. Then through trial and error (evaluating the formula for $\mathcal{L}_{\mathrm{MS}}$ shown in Section 3, for a series of increasingly unbalanced MSs) it can be shown that, for $K=10$ (say) and $\alpha=0.05$, if the shares are (0.01662, 0.98338) or more unbalanced (a situation that is not unreasonable in many applications), the critical MS-based ρ^2 is greater than 1, which is impossible to achieve, meaning that it will be impossible to find a model that is *statistically superior* (based on the χ^2 likelihood ratio test) to the MS model. While that is useful to know, as discussed in Section 3 it should not deter us from trying to explain *why* the MSs are so unbalanced, through the inclusion of real variables into the model.

The ρ^2 measure has fallen out of favor with some scholars, who prefer to focus on the χ^2 likelihood ratio test for comparing nested models. One virtue of the ρ^2 measure, however, lies in its ability to provide a sense of how good a model is in its own right, not just in comparison to other models. Many very elaborate models have "barely moved the log-likelihood needle" off that of the benchmark model (whichever one is chosen), but the discussion focuses exclusively on the statistical significance of individual parameter estimates, and on the "significant" (but tiny) improvement of the full model over the benchmark model, without addressing the big picture of how much information is still *not* explained by the full model. In my view, the latter is worth keeping in mind, much as we would generally consider reporting an R^2 measure to be *de rigueur* for a regression model.

6. Computation and interpretation of ρ^2 under certain special circumstances

There are several common circumstances in which the standard formula of Eq. (1) needs refinement. I briefly mention three such situations below.

6.1. Repeated choices with no change in observed explanatory variables

Often, the analyst has frequency or share data representing an individual's distribution of choices among multiple alternatives over time, rather than discrete data representing the choice on a single occasion. Hauser (1978, p. 409) identifies this as one case in which "$\rho^2=1$ may not be the appropriate upper bound. If individuals make repeated choices and do not always select the same alternative [given the same observed explanatory variables], then $\rho^2=1$ is not possible, even in theory. (Perfect prediction would require different probabilities for different occasions. Such predictions are not possible without situational variables)." As he indicates, in this instance, the best possible model is one for which the predicted probability of choice is always equal to the observed relative frequency of choice, say f_{jn} for person n choosing alternative j. In that case, the theoretical maximum log-likelihood is

$$\mathcal{L}_{\max} = \sum_n \sum_j f_{jn} \ln f_{jn} \tag{8}$$

rather than 0 (see Mokhtarian and Bagley, 2000 for an empirical application).

6.2. Individual-specific choice sets

Section 6.1 indicated a situation in which the *upper* bound of $\mathcal{L}$ is different than in the standard case; here we consider a situation in which its *lower* bound is different. When choice sets differ by individual, $\mathcal{L}_{\mathrm{EL}}$ is not $-N \ln J$ as it is when all N cases have the same J alternatives, but rather $-\Sigma_n \ln J_n$, where $J_n \le J$ is the number of alternatives in person n's choice set. If $J_n < J$ for any n, then this number will be larger (less negative) than $-N \ln J$, reflecting the information contained in the assumption that some alternatives have a zero probability of being selected rather than the equally-likely (maximum-entropy, zero-information) probability of $1/J$ (see Hauser, 1978 for an example). A similar comment applies to the market-share log-likelihood, $\mathcal{L}_{\mathrm{MS}}$. Thus, the computation of ρ^2 must take that adjustment into account.

6.3. Market segmentation

Suppose the market is divided into G segments deterministically based on observed characteristics (for example, high-, medium-, and low-income), with coefficients $\boldsymbol{\beta}^g$ allowed to differ by segment. It is useful to be able to compare the ρ^2 for the segmented model to that for the pooled model, in which the coefficients for a given variable are constrained to be equal across segment. If viewed as a single "giant" model in which all coefficients are segment-specific, it is easy to show (as given in Ben-Akiva and Lerman 1985, p. 195) that the $\mathcal{L}$ for the segmented model is the sum of the $\mathcal{L}(\hat{\boldsymbol{\beta}}^g)$s for the models on each segment separately (as long as specifications and samples do not change). In that case, it is also easy to see that the appropriate EL-base ρ^2 for the segmented model is:

$$\textit{EL-base}\ \rho^2_{\text{segmented}} = 1 - \left[\sum_g \mathcal{L}(\hat{\boldsymbol{\beta}}^g) / \sum_g \mathcal{L}(\mathbf{0}^g) \right], \tag{9}$$

where N_g is the number of cases in the gth segment, $g = 1, 2, \ldots, G$, and (when choice sets are equal)

$$\sum_g \mathcal{L}(\mathbf{0}^g) = \sum_g (-N_g \ln J) = -\ln J \sum_g N_g = -N \ln J, \tag{10}$$

meaning that $\mathcal{L}_{\mathrm{EL}}$s for the segmented and pooled models are equal when choice sets are equal. Note, however, that the same is not true for the MS model. So to compute $\rho^2_{\text{segmented}}$ *for the MS base*, the denominator cannot be the pooled $\mathcal{L}_{\mathrm{MS}}$, but rather must be the segmented version, namely $\Sigma_g \mathcal{L}^g_{MS}$.

Acknowledgments

Comments by two anonymous referees have prompted numerous improvements to this article.

References

Ben-Akiva, Moshe, Lerman, Steven R., 1985. In: Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA.
Ben-Akiva, Moshe, Swait, Joffre, 1986. The Akaike likelihood ratio index. Transp. Sci. 20 (2), 133–136.
Cameron, A. Colin, Windmeijer, Frank A.G., 1997. An *R*-squared measure of goodness of fit for some common nonlinear regression models. J. Economet. 77, 329–342.
Domencich, Thomas A., McFadden, Daniel, 1975. In: Urban Travel Demand: A Behavioral Analysis. North-Holland Publishing Co., Amsterdam, Available at ⟨http://elsa.berkeley.edu/users/mcfadden/travel.html⟩ (accessed 02.08.15).
Greene, William H., 2009. In: NLOGIT 5 Reference Guide. Econometric Software, Inc., Plainview, NY.
Hauser, John R., 1978. Testing the accuracy, usefulness, and significance of probabilistic choice models: an information-theoretic approach. Oper. Res. 26 (3), 406–421.
Hensher, David A., Rose, John M., Greene, William H., 2005. In: Applied Choice Analysis: A Primer. Cambridge University Press, Cambridge, UK.
Manski, Charles F., Lerman, Steven R., 1977. The estimation of choice probabilities from choice-based samples. Econometrica 45, 1977–1988.
McFadden, Daniel, 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics, Academic Press, New York, pp. 105–142. (Chapter 4).
Menard, Scott, 2000. Coefficients of determination for multiple logistic regression analysis. Am. Stat. 54 (1), 17–24.
Mokhtarian, Patricia L., Bagley, Michael N., 2000. Modeling employees' perceptions and proportional preferences of work locations: the regular workplace and telecommuting alternatives. Transp. Res. A 34, 223–242.
Ortúzar, Juan de Dios, Willumsen, Luis G., 2011. In: Modelling Transport, 4th ed. John Wiley and Sons, Chichester, UK.
Tardiff, Timothy J., 1976. A note on goodness-of-fit statistics for probit and logit models. Transportation 5, 377–388.
Train, Kenneth, 2009. In: Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, UK, Available at ⟨http://elsa.berkeley.edu/~train/distant.html⟩ (accessed August 2, 2015).
Veall, Michael R., Zimmermann, Klaus F., 1996. Pseudo-R^2 measures for some common limited dependent variable models. J. Econ. Surv. 10 (3), 241–259.
Windmeijer, Frank A.G., 1995. Goodness-of-fit measures in binary choice models. Econ. Rev. 14 (1), 101–116.