UNSW THE UNIVERSITY OF NEW SOUTH WALES

Australian School of Business

Working Paper

Never Stand Still Australian School of Business

Australian School of Business Research Paper No. 2012 ECON 48

# A new condition for pooling states in multinomial logit

Hong il Yoo

# A new condition for pooling states in multinomial logit

Hong il Yoo[a,*]

[a] School of Economics, University of New South Wales, Australia. h.yoo@unsw.edu.au

October 24, 2012

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

**Abstract**

The Cramer-Ridder test is a popular procedure for testing if some outcome states can be pooled into one state in the multinomial logit model. We show that, in the presence of binary regressors, the test is overly stringent and poolability may not be tested unambiguously.

**JEL classification:** C35
**Key words:** multinomial logit, pooling, statistical test

**Highlights:**

- We revisit the problem of pooling states in the multinomial logit model.
- The pooling condition and test due to Cramer and Ridder (1991) are well-known.
- This condition is at odds with how saturated models behave.
- We derive a new condition for pooling which is not.
- With several binary regressors, pooling may not be unambiguously tested.

# 1 Introduction

Empirical researchers often face decisions over pooling unordered categorical outcomes. When modelling individual labour force status, for example, non-participation may be treated as one homogeneous state or a further distinction may be drawn between those who are not desiring work and those who are but not searching actively.

Cramer and Ridder (1991) have developed a well-known procedure for testing if a subset of categorical outcomes, or states, can be pooled into one state in the multinomial logit (MNL) model. The authors show that separate states can be viewed as arbitrary subdivisions of one parent state when their slope coefficients are identical, and pooled without affecting the probabilities of the other states. The Cramer-Ridder test, or likelihood ratio test of the corresponding parametric restrictions, is often applied in practice (Chalkley and McVicar, 2008; Dancer and Fiebig, 2004; Ngyuen and Taylor, 2003) and also available as a Stata command, -crtest-.

We show that, in the presence of binary regressor(s), not all slope coefficients need be identical across states to make their distinction irrelevant *à la* Cramer and Ridder (1991). A sufficient condition for pooling can be derived from a subdivision process which nests the one Cramer and Ridder have postulated. This condition allows for state-specific coefficients on those binary regressors which collectively permit saturated coefficient parameterisations; it only imposes the cross-state equality of coefficients on the other regressors. A corresponding test of pooling is a test of the cross-state equality of coefficients on continuous regressors when there is only one binary regressor. With several binary regressors, a test of pooling may not be implemented as unambiguously as Cramer and Ridder have prescribed; several pooling conditions may coexist, each implying a different set of parametric restrictions.

# 2 Pooling states in MNL

Cramer and Ridder (1991; C&R hereafter) analyse an MNL model with $S$ mutually exclusive states to establish when the distinction of separate states may be deemed irrelevant. Using $R_n$ to denote the sum $1 + \sum_{j=2}^{S-1} \exp(\alpha_j + \boldsymbol{\delta}_j \cdot \mathbf{d}_n + \boldsymbol{\gamma}_j \cdot \mathbf{g}_n + \boldsymbol{\beta}_j \cdot \mathbf{x}_n)$, the probability of state $s$ at observation $n$ is specified as:

$$P_{ns} = \frac{\exp(\alpha_s + \boldsymbol{\delta}_s \cdot \mathbf{d}_n + \boldsymbol{\gamma}_s \cdot \mathbf{g}_n + \boldsymbol{\beta}_s \cdot \mathbf{x}_n)}{R_n + \exp(\alpha_S + \boldsymbol{\delta}_S \cdot \mathbf{d}_n + \boldsymbol{\gamma}_S \cdot \mathbf{g}_n + \boldsymbol{\beta}_S \cdot \mathbf{x}_n)} \tag{1}$$

given continuous regressors $\mathbf{x}_n$, and two sets of binary regressors $\mathbf{d}_n$ and $\mathbf{g}_n$ distinguished as follows. Had $\mathbf{d}_n$ been the only regressors, the model would be saturated for $\mathbf{d}_n$; the same statement does not apply to $\mathbf{g}_n$. Such division of the regressors has not been considered by C&R, but plays a major role in our analysis. $\alpha_s$ is the intercept, $\boldsymbol{\delta}_s, \boldsymbol{\gamma}_s$ and $\boldsymbol{\beta}_s$ are the slope coefficients associated with state $s = 1, 2, \cdots, S$; all these parameters are normalised to 0 in state 1.

C&R then suppose that state $S$ is randomly split into $Q$ superficially different states, $S_1, S_2, \cdots, S_Q$, with the probability $\lambda_{S_q}$ of subdivision into $S_q$ so that $P_{nS_q} = \lambda_{S_q} P_{nS}$ and $P_{nS} = \sum_{q=1}^{Q} P_{nS_q}$. The resulting model with $S-1+Q$ states still retains the MNL form, with $P_{ns}$ unchanged from (1) for any state $s = 1, 2, \cdots, S-1$, and $P_{nS_q}$ for each new state $S_q$ being:

$$
\begin{aligned}
P_{nS_q} &= \frac{\exp(\alpha_{S_q} + \boldsymbol{\delta}_S \cdot \mathbf{d}_n + \boldsymbol{\gamma}_S \cdot \mathbf{g}_n + \boldsymbol{\beta}_S \cdot \mathbf{x}_n)}{R_n + \sum_{l=1}^{Q} \exp(\alpha_{S_l} + \boldsymbol{\delta}_S \cdot \mathbf{d}_n + \boldsymbol{\gamma}_S \cdot \mathbf{g}_n + \boldsymbol{\beta}_S \cdot \mathbf{x}_n)} \\
&= \frac{\exp(\alpha_{S_q} + \boldsymbol{\delta}_S \cdot \mathbf{d}_n + \boldsymbol{\gamma}_S \cdot \mathbf{g}_n + \boldsymbol{\beta}_S \cdot \mathbf{x}_n)}{R_n + \exp(\alpha_S + \boldsymbol{\delta}_S \cdot \mathbf{d}_n + \boldsymbol{\gamma}_S \cdot \mathbf{g}_n + \boldsymbol{\beta}_S \cdot \mathbf{x}_n)}
\end{aligned} \tag{2}
$$

where $\alpha_{S_q} = \alpha_S + \ln \lambda_{S_q}$. Note that only the intercept varies across the $Q$ new states.

C&R conclude that conversely, $Q$ out of the $S$ states in the original model, (1), can be pooled into one state when the $Q$ states share the same slope coefficients, because then they can be viewed as subdivisions of one parent state. With $\Psi$ denoting the set of the $Q$ candidate states, the likelihood ratio test of these parametric restrictions:

$$
\boldsymbol{\delta}_j = \boldsymbol{\delta}_k, \boldsymbol{\gamma}_j = \boldsymbol{\gamma}_k, \boldsymbol{\beta}_j = \boldsymbol{\beta}_k \text{ for all } j, k \in \Psi, j \neq k \tag{3}
$$

is known as the Cramer-Ridder test (C-R test hereafter). Each set of restrictions can be ignored when the corresponding type of regressor is not present.

To see that the pooling condition (3) may be overly stringent, consider a simple MNL model wherein $\mathbf{d}_n$ are the only regressors.[1] The ML estimator of each possible $P_{ns}$ equals the sample relative frequency of state $s$ among observations described by the same configuration of $\mathbf{d}_n$, consistently estimating the corresponding population frequency. Accordingly, any subset of the other states can be merged without affecting substantive results regarding state $s$, even though $\boldsymbol{\delta}_j$ may vary across those states.

[1] For example, consider the following three cases: (a) there is only one regressor and that regressor is binary (b) the only regressors are binary indicators of different categories of the same qualitative variable (eg. region of residence) and (c) the only regressors are such binary indicators of several qualitative variables and all cross-products of those indicators.

We now derive from the same $S$-state model, (1), a new pooling condition compatible with the algebraic property of the ML estimator. As before, state $S$ is randomly subdivided into $Q$ new states, $S_1, S_2, \cdots, S_Q$, but we generalise the subdivision probabilities to vary with $\mathbf{d}_n$, instead of being the same for every $n$. The probability of subdivision into $S_q$ is denoted $f_{S_q}(\mathbf{d}_n)$ to emphasise its dependence on $\mathbf{d}_n$, but otherwise plays an analogous role as $\lambda_{S_q}$ above. $P_{ns}$ remains unchanged from (1) for any state $s = 1, 2, \cdots, S-1$, while $P_{nS_q}$ for each new state $S_q$ is:

$$\begin{aligned} P_{nS_q} &= \frac{\exp(\alpha_S + \ln f_{S_q}(\mathbf{d}_n) + \boldsymbol{\delta}_S \cdot \mathbf{d}_n + \boldsymbol{\gamma}_S \cdot \mathbf{g}_n + \boldsymbol{\beta}_S \cdot \mathbf{x}_n)}{R_n + \sum_{l=1}^{Q} \exp(\alpha_S + \ln f_{S_l}(\mathbf{d}_n) + \boldsymbol{\delta}_S \cdot \mathbf{d}_n + \boldsymbol{\gamma}_S \cdot \mathbf{g}_n + \boldsymbol{\beta}_S \cdot \mathbf{x}_n)} \\ &= \frac{\exp(\alpha_S + \ln f_{S_q}(\mathbf{d}_n) + \boldsymbol{\delta}_S \cdot \mathbf{d}_n + \boldsymbol{\gamma}_S \cdot \mathbf{g}_n + \boldsymbol{\beta}_S \cdot \mathbf{x}_n)}{R_n + \exp(\alpha_S + \boldsymbol{\delta}_S \cdot \mathbf{d}_n + \boldsymbol{\gamma}_S \cdot \mathbf{g}_n + \boldsymbol{\beta}_S \cdot \mathbf{x}_n)} \end{aligned} \tag{4}$$

The new model with $S-1+Q$ states seems to have a different regressor specification from the parent $S$-state model, due to $\ln f_{S_q}(\mathbf{d}_n)$ terms in states $S_1$ through $S_Q$. But when there exist real numbers $\{a_{S_q}, \mathbf{b}_{Sq}\}_{q=1}^{Q}$ such that:

$$\begin{aligned} \ln f_{S_q}(\mathbf{d}_n) &= a_{S_q} + \mathbf{b}_{S_q} \cdot \mathbf{d}_n \text{ for } q = 1, \cdots, Q-1 \\ \ln f_{S_Q}(\mathbf{d}_n) &= \ln(1 - \textstyle\sum_{q=1}^{Q-1} f_{S_q}(\mathbf{d}_n)) = a_{S_Q} + \mathbf{b}_{S_Q} \cdot \mathbf{d}_n \end{aligned} \tag{5}$$

for all configurations of $\mathbf{d}_n$, the new model is observationally equivalent to an MNL model with $S-1+Q$ states using the same specification as the parent model. In that equivalent model, obtained by plugging (5) into (4), the intercept varies across the new states, $\alpha_{S_q} = \alpha_S + a_{S_q}$, while the coefficients on $\mathbf{g}_n$ and $\mathbf{x}_n$ do not, $\boldsymbol{\gamma}_{Sq} = \boldsymbol{\gamma}_S$ and $\boldsymbol{\beta}_{Sq} = \boldsymbol{\beta}_S$, where $q = 1, \cdots, Q$. But unlike what the C-R test imposes under the null, (3), the coefficients on $\mathbf{d}_n$ vary across the new states too, $\boldsymbol{\delta}_{Sq} = \boldsymbol{\delta}_S + \mathbf{b}_{S_q}$. As can be easily seen, the existence of $\{a_{S_q}, \mathbf{b}_{Sq}\}_{q=1}^{Q}$ as in (5) is also necessary for that of such equivalent model.

Because the subdivision probabilities are saturated for $\mathbf{d}_n$, the linear system (5) always exists, regardless of the functional form of $\{f_{S_q}(\mathbf{d}_n)\}_{q=1}^{Q}$. In the context of the original $S$-state model, the pooling condition replacing (3) is:

$$\boldsymbol{\gamma}_j = \boldsymbol{\gamma}_k, \boldsymbol{\beta}_j = \boldsymbol{\beta}_k \text{ for all } j, k \in \Psi, j \neq k \tag{6}$$

Either set of restrictions can be ignored when the corresponding type of regressor is not present. (6) originates from a subdivision process which nests the one C&R have postulated for (3). Moreover, (6) leaves the coefficients on $\mathbf{d}_n$ unrestricted, in agreement with how saturated MNL models behave.

No such pooling condition arises when the subdivision probabilities are generalised further to vary also with $\mathbf{g}_n$ and/or $\mathbf{x}_n$. Suppose first that they vary with both $\mathbf{d}_n$ and $\mathbf{g}_n$, and are denoted $\{h_{S_q}(\mathbf{d}_n, \mathbf{g}_n)\}_{q=1}^{Q}$. To permit pooling, there must be real numbers $\{a_{S_q}, \mathbf{b}_{Sq,}\mathbf{c}_{S_q}\}_{q=1}^{Q}$ such that:

$$\begin{aligned} \ln h_{S_q}(\mathbf{d}_n, \mathbf{g}_n) = a_{S_q} + \mathbf{b}_{S_q} \cdot \mathbf{d}_n + \mathbf{c}_{S_q} \cdot \mathbf{g}_n \text{ for } q = 1, \cdots, Q-1 \\ \ln h_{S_Q}(\mathbf{d}_n, \mathbf{g}_n) = \ln(1 - \textstyle\sum_{q=1}^{Q-1} h_{S_q}(\mathbf{d}_n, \mathbf{g}_n)) = a_{S_Q} + \mathbf{b}_{S_Q} \cdot \mathbf{d}_n + \mathbf{c}_{S_Q} \cdot \mathbf{g}_n \end{aligned} \tag{7}$$

for all configurations of $\mathbf{d}_n$ and $\mathbf{g}_n$. Because $\mathbf{d}_n$ and $\mathbf{g}_n$ collectively do not permit saturated linear parameterisations, (7) may not be satisfied for all functional forms of $\{h_{S_q}(\mathbf{d}_n, \mathbf{g}_n)\}_{q=1}^{Q}$; a given variation in some binary regressors may shift a particular subdivision probability by different proportions depending on configurations of all binary regressors. To satisfy (7), each $h_{S_q}(\mathbf{d}_n, \mathbf{g}_n)$ needs to be a product of the parts that vary with $\mathbf{d}_n$ only, $h^A_{S_q}(\mathbf{d}_n)$, and with $\mathbf{g}_n$ only, $h^B_{S_q}(\mathbf{g}_n)$. Such multiplicative decomposition, however, is not possible because $h_{S_Q}(\mathbf{d}_n, \mathbf{g}_n) = 1 - \sum_{q=1}^{Q-1} h_{S_q}(\mathbf{d}_n, \mathbf{g}_n)$; for two different configurations of $\mathbf{g}_n$, $\mathbf{g}^0_n$ and $\mathbf{g}^1_n$, the ratio $h_{S_Q}(\mathbf{d}_n, \mathbf{g}^1_n)/h_{S_Q}(\mathbf{d}_n, \mathbf{g}^0_n)$ must depend on both $\mathbf{d}_n$ and $\mathbf{g}_n$, even when $h_{S_q}(\mathbf{d}_n, \mathbf{g}_n) = h^A_{S_q}(\mathbf{d}_n)h^B_{S_q}(\mathbf{g}_n)$ for all $q < Q$. An analogous argument establishes that the subdivision probabilities also varying with $\mathbf{x}_n$ do not lead to a pooling condition.

# 3 Practical implications

When there are only continuous regressors $\mathbf{x}_n$, the pooling condition (3) due to C&R and ours (6) coincide.

When at least one binary regressor is present, (3) is stronger than (6) because any binary regressor, when viewed in isolation, satisfies the definition of $\mathbf{d}_n$. The poolability of the candidate states may thus be incorrectly rejected according to the C-R test because (3) is sufficient but not necessary for pooling. In particular, when both binary and continuous regressors are present but all binary regressors can be classified as $\mathbf{d}_n$, the C-R test imposes the cross-state equality of all slope coefficients under the null, while to test (6) is to test that of the continuous regressor coefficients alone.[2]

Under more general settings, a test of pooling in the MNL model may not be "of almost trivial simplicity" as Cramer and Ridder (1991) have concluded.

For illustration, suppose that there are continuous regressors, a set of dummies indicating different regions of residence, and another set of dummies indicating different levels of self-assessed health. Each set of dummies define mutually exclusive groups of observations, but the two sets together do not permit saturated coefficient parameterisations because self-assessed health may vary within each region. (6) may thus be tested in two major forms, classifying the regional dummies as $\mathbf{d}_n$ and the health dummies as $\mathbf{g}_n$ or vice versa. These two implementations hypothesise different data generating processes and need not result in the same conclusion.

Such ambiguity arises because, given a selection of regressors, more than one subdivision processes may be invoked to justify pooling the candidate states. To test pooling, then, becomes to test restrictions implied by a maintained subdivision process. The C-R test always prescribes an unambiguous procedure only because it maintains a particular subdivision process which is restrictive and has implications at odds with how saturated MNL models behave.

In light of (6), arguably the only unambiguous step of a possible testing strategy would be to start with a test of the equality of the continuous regressor coefficients:

$$\boldsymbol{\beta}_j = \boldsymbol{\beta}_k \text{ for all } j, k \in \Psi, , j \neq k \tag{8}$$

which is both sufficient and necessary for pooling when only $\mathbf{d}_n$ and $\mathbf{x}_n$ are present, and necessary when $\mathbf{g}_n$ are present too. When the null is not rejected in the latter case, further tests of pooling may require more subjective decisions over which of subdivision processes to maintain, either because testing (6) based on each possible classification of $\mathbf{d}_n$ and $\mathbf{g}_n$ leads to conflicting conclusions, or because it is impractical and/or uninteresting to compute test statistics for all such classifications.

[2]For example, suppose that each special case in footnote 1 is extended by including continuous regressor(s). The empirical example of Cramer and Ridder (1991) corresponds to the first of the resulting cases because it includes one binary regressor and one continuous regressor.

## Acknowledgements

Helpful discussion with Denzil G. Fiebig is gratefully acknowledged. This work was supported under Australian Research Council's Discovery Projects funding scheme (project number: DP0881205), without any conflict of interest.

## References

Chalkley, M., McVicar, D., 2008. Choice of contracts in the British National Health Service: An empirical study. Journal of Health Economics. 27, 1155-1167.

Cramer, J.S., Ridder, G., 1991. Pooling states in the multinomial logit model. Journal of Econometrics. 47, 267-272.

Dancer, D.M., Fiebig, D.G., 2004. Modelling students at risk. Australian Economic Papers. 43, 158-173.

Ngyuen, A.N., Taylor, J., 2003. Post-high school choices: New evidence from a multinomial logit model. Journal of Population Economics. 16, 287-306.