Multinomial Logit Normal Temporal Model

Carolina Franco

1 Approximately incorporating the effect of complex sampling design in multinomial models.

We now extend the setting of §?? so that each of the N_i elements in the ith small area takes on one of K categories. The finite population count for the kth category in the ith small area is N_{ik}, $\sum_{k=1}^{K} N_{ik} = N_i$, and $N_{ik}/N_i = p_{ik}$. In much of our discussion, we will consider N_i large so that p_{ik} can be treated as the success probability for category k in domain i. Let $\{\widehat{p}_{ik}\}_{i=1}^{m}$ denote survey-weighted direct estimates of these small area proportions, let $v_{ik} = \mathrm{Var}(\widehat{p}_{ik})$ denote the design variance of the direct estimate and $\widehat{v}_{ik}$ an estimator of v_{ik}. The former estimator should be smoothed, using the techniques studied in §??. As with the binomial-type models described in §??, one must approximately account for the sampling design. However, in the multinomial case, one must take into account that the design effect might differ among multinomial categories. We first define the (category-specific) effective sample sizes and counts by analogy to (??)–(??):

$$\tilde{n}_{ik} = \tilde{p}_{ik}(1 - \tilde{p}_{ik}) \,/ \widehat{v}_{ik}, \tag{1}$$

$$\tilde{y}_{ik} = \tilde{n}_{ik} \times \widehat{p}_{ik}. \tag{2}$$

We then propose to define the multinomial effective sample size as the sum of the effective sample counts (2) over all the categories:

$$\tilde{n}_i = \sum_{k=1}^{K} \tilde{y}_{ik} = \sum_{k=1}^{K} \tilde{n}_{ik} \widehat{p}_{ik}. \tag{3}$$

Note that (3) is a convex combination of the marginal effective sample sizes over all the categories, weighted by their corresponding proportions.

An alternative proposal for the multinomial effective sample size is given in McAllister and Ianelli (1997) and used by Maples (2019). It is also a convex combination of effective sample sizes, but with different weights:

$$\tilde{n}_i^* = \sum_{k=1}^{K} \tilde{n}_{ik} \left(\frac{\widehat{v}_{ik}}{\sum_{\ell=1}^{K} \widehat{v}_{i\ell}} \right).$$

Both of these proposals reduce to the usual binomial effective sample size for $K = 2$, as would be true for any convex combination of the effective sample sizes.

A third alternative would be to use the representation of the multinomial distribution as a product of binomials (see Linderman et al. (2015), formulas (5) and (6)). Each binomial would then use its corresponding usual effective sample size.

The appropriate way of handling the effective sample size in a multinomial for a small area model is another area of practical importance that has received limited attention in the literature. We will evaluate these alternative proposals via simulation, analytical theory, and empirical study.

2 Temporal multinomial logit normal model.

Often, predictive covariates are hard to find for categorical outcomes, but estimates are available for various time points in a survey. In that case, a promising approach is to borrow strength over time, while capturing the sampling dependence of all the survey estimates used across the years. We propose to do this using VAR(1) extensions of multinomial SAE models using the Bayesian paradigm. In preliminary work, we have successfully tested the model proposed below with $K = 4$ categories using custom STAN code. We describe the model with $K = 4$ for simplicity.

Let $\tilde{y}_{ik,t}$ denote the effective sample counts for domain i and category k at time t and let $\tilde{n}_{i,t}$ denote the associated effective sample size for domain i at time t, following one of the approaches described in §1. Define

$$\boldsymbol{p}_{i,t} = (p_{i1,t}, \ldots, p_{iK,t})^\top = \left(\frac{N_{i1,t}}{N_{i,t}}, \ldots, \frac{N_{iK,t}}{N_{i,t}} \right)^\top .$$

Then, the temporal Multinomial Logit Normal (MLN) model is

$$(\tilde{y}_{i1,t}, \tilde{y}_{i2,t}, \tilde{y}_{i3,t}, \tilde{y}_{i4,t})^\top \mid \tilde{n}_{i,t}, \boldsymbol{p}_{i,t} \sim \text{Multinom}\left(\tilde{n}_{i,t}, \boldsymbol{p}_{i,t}\right). \tag{4}$$

where

$$\log(p_{ik,t}/p_{i4,t}) = \mathbf{x}_{ik,t}^\top \boldsymbol{\beta}_{k,t} + u_{ik,t} \quad k = 1, 2, 3; i = 1, 2, \ldots, m; \tag{5}$$

$$\mathbf{u}_{i,t} = \mathbf{A}_t \mathbf{u}_{i,(t-1)} + \boldsymbol{\epsilon}_{i,t}; \tag{6}$$

$\mathbf{A}_t$ is a 3×3 matrix; $\{\boldsymbol{\epsilon}_{i,t}\}$ is iid trivariate $\mathcal{N}(\mathbf{0}, \Sigma_\epsilon)$; and $\mathbf{u}_{i,t} = (u_{i1,t}, u_{i2,t}, u_{i3,t})^\top$.

An alternative parameterization that has shown better behavior in preliminary work expresses the multinomial parameters $\boldsymbol{p}_{i,t}$ in terms of conditional probabilities,

$$\boldsymbol{p}_{i,t} = (p_{i1,t}, \ldots, p_{iK,t})^\top = ((1 - \mu_{i,t}),\ \mu_{i,t}(1 - \nu_{i,t}),\ \mu_{i,t}\nu_{i,t}(1 - \rho_{i,t}),\ \mu_{i,t}\nu_{i,t}\rho_{i,t})^\top . \tag{7}$$

This parameterization is inspired by Slud et al. (2018), which uses a Dirichlet-multinomial distribution in an application involving nested categories to inform the enforcement of the Voting Rights Act (VRA), Section 203b, in 2016 (Slud et al., 2018). With this parameterization, another version of the temporal MLN model can then be defined with the sampling model (4) and with the baseline category logits (5) replaced by

$$\text{logit}(\mu_{i,t}) = \mathbf{x}_{i1,t}\boldsymbol{\beta}_{1,t} + u_{i1,t},\ \text{logit}(\nu_{i,t}) = \mathbf{x}_{i2,t}\boldsymbol{\beta}_{2,t} + u_{i2,t},\ \text{logit}(\rho_{i,t}) = \mathbf{x}_{i3,t}\boldsymbol{\beta}_{3,t} + u_{i3,t},$$

and with the vector autoregression specified as before in (6).

Of course, one could use the probit transformation instead of the logit above, among other alternatives. An early paper to apply multinomial logit normal models to SAE was Molina et al. (2007). Their model, however included a single random effect for all multinomial categories, a rather strong assumption. López-Vizcaíno et al. (2013) did have separate random effects, but they were assumed to be independent. López-Vizcaíno et al. (2015) is a temporal extension of the former, but with some limitations. This work, handled from a frequentist approach, assumes both independence of the random effects across multinomial categories, and an independent univariate AR(1) structure for each category. These independence assumptions are unlikely to hold in many real applications. Our proposed model is a more general case, and permits for hypothesis tests of the independence assumptions. It is most appropriate for applications with a rather large number of areas, but these are quite common given the growing demand for disaggregation.

Furthermore, the methodology of López-Vizcaíno et al. (2015) was developed and tested for the case of three categories, and further development into more categories may be computationally challenging using their framework. In contrast, the model we propose has already been shown in preliminary work to work for 4 categories, and we suspect it will work well with more categories as well. Real problems with many categories abound in practice.

Among other applications, this model has the potential for being useful for the Voting Rights Act, Section 203. The non-temporal version of this model was proposed by the PI of this proposal for the 2021 determinations when she led the model development for that program in her former role as Small Area Estimation Group Leader in the Center of Statistical Research and Methodology at the US Census Bureau, and the model was used in official production. The VRA, Section 203 statistical program is extremely important as it used to enforce which parts of the United States are required by law to print voting materials in other languages for various Language Minority Groups. The preliminary work that she has done has been using data on this problem. In past iterations of the official production of VRA, Section 203, covariates that were used in this application were drawn from the same survey as the response, but such covariates would be subject to measurement error, which can create problems (Bell et al., 2019). Furthermore, if there is overlap in the sample with the response, a misspecification in the assumed independence between the model error and the sampling error occurs. The proposed model borrows strength over time instead, while capturing the relevant uncertainties and dependencies.

References

Bell, W. R., H. C. Chung, G. S. Datta, and C. Franco (2019). Measurement error in small area estimation: Functional versus structural versus naïve models. *Survey Methodology 45*(1), 61–80.

Linderman, S., M. J. Johnson, and R. P. Adams (2015). Dependent multinomial models made easy: Stick-breaking with the Pólya-gamma augmentation. *Advances in Neural Information Processing Systems 28*.

López-Vizcaíno, E., M. J. Lombardía, and D. Morales (2013). Multinomial-based small area estimation of labour force indicators. *Statistical modelling 13*(2), 153–178.

López-Vizcaíno, E., M. J. Lombardía, and D. Morales (2015). Small area estimation of labour force indicators under a multinomial model with correlated time and area effects. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(3), 535–565.

Maples, J. (2019). Small area estimates of the child population and poverty in school districts using Dirichlet-multinomial models. In *Proceedings of the American Statistical Association, Survey Research Methods Section*, pp. 2448–2456.

McAllister, M. K. and J. N. Ianelli (1997). Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. *Canadian Journal of Fisheries and Aquatic Sciences 54*(2), 284–300.

Molina, I., A. Saei, and M. Jose Lombardia (2007). Small area estimates of labour force participation under a multinomial logit mixed model. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 170*(4), 975–1000.

Slud, E., R. Ashmead, P. Joyce, and T. Wright (2018). Statistical methodology (2016) for Voting Rights Act, Section 203 determinations. *Statistics*, 12.