Should small samples from recent time point be used with older data? Applicability of updating models by transfer scaling

Nobuhiro Sanko

School of Business Administration, Kobe University

Abstract:

[Background]
Use of recent data is crucial for disaggregate travel demand forecasting. There is a significant body of evidence that forecasts by models with the recent data outperform those by models with older data. However, enough samples are not always available from the recent time point. Therefore, an attempt has been made to collect smaller samples and use them together with older data, which is expected to result in better forecasts than using only the older data. A model updating is a technique to update models estimated with greater number of observations from the older time point by utilising smaller number of observations from the recent time point. One such technique is transfer scaling, where models are estimated with the older data and only alternative-specific constants and scale parameters of utility functions are updated by using the recent data, while other parameters related to level-of-service and socio-economic variables are used as is.

Although model updating utilises data from two points in time, there is a case that a use of only smaller samples from the recent time point results in better forecasts than using both datasets. Badoe and Miller (1995) and Karasmaa and Pursula (1997) argue that even when the sample size of the recent data is small, say around 400-500, additional use of the older data contributes little to improving forecasting.

However, these studies have the following limitations:
1. The sample size of the older data is fixed, and only the sample size of the recent data is varied.
2. Smaller samples from the recent time point is chosen on a random basis, but their random draw was done only once, thereby results being coincidental.
3. Data are available from only two points in time, and the data from the second time point was used both for model updating and model validation.

The present study examines: should small samples from recent time point be used with older data? More specifically, in which case should only small samples collected recently be used? In which case should transfer scaling be used by utilising both datasets? The author answers this by utilising repeated cross-sectional data collected at four points in time. The sample sizes of the recent and older data are varied simultaneously. Furthermore, a bootstrapping technique is adopted to find results with statistical meaning.

[Data]
The repeated cross-sectional data used in this study come from household travel surveys implemented in

the Nagoya Metropolitan Area, Japan, in 1971, 1981, 1991, and 2001. This study utilises data collected at the three time points of 1971, 1981, and 1991 for forecasting travel behaviour in 2001. The trips modelled are journeys to work (commuting) mode choice. Transport modes considered are rail, bus, and car.

[Methodology]
Five notations are introduced: y1, y2, n1, n2, and b. The y1 and y2 denote years when the data was collected: 1971, 1981, and 1991 (y1 < y2). The n1 and n2 represent the sample size from y1 and y2, respectively (n1 >= n2). The author examined 12 n's: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, and 10000. The b (= 1, 2, . . ., 200) represents a bootstrap repetition. The two inequalities in the above two parentheses mean that models estimated using larger samples (n1) from the older time point (y1) are updated by using smaller samples (n2) collected recently (y2).

For each y1, y2, n1, and n2, samples are randomly drawn 200 (b = 1, 2, . . ., 200) times with replacement. In total, 3 combinations of y's ((y1, y2) = (1971, 1981), (1971, 1991), and (1981, 1991)) * 78 combinations of n's (12*13/2) * 200 b's = 46,800 datasets are generated. For each y1, y2, n1, n2, and b, (1) a multinomial logit model is estimated by using n1 observations from y1, and the model is updated by using n2 observations from y2; (2) a multinomial logit model is estimated by using n2 observations from y2. These two types of models are applied to forecast behaviours of 2001.

The forecast performance is evaluated by log-likelihood on the 2001 dataset. Let L1 (y1, y2, n1, n2, b) and L2 (y2, n2, b) represent log-likelihoods by the above models (1) and (2), respectively. Note that L1 and L2 are not always available since smaller n1 and/or n2 sometimes cause poor estimates.

In addition, the following variable xb is defined for b = 1, 2, …, 200 only when both L1 and L2 are available.

xb = L1 (y1, y2, n1, n2, b) - L2 (y2, n2, b)

The author assumes that xb is normally distributed, and statistical tests are conducted.

[Results]
Based on statistical tests, in any combinations of sample sizes from two time points, transfer scaling never produces statistically significantly better forecasts than using only small samples from the recent time point. On the other hand, when the sample size from the recent time point increases, using only the most recent data produces statistically significantly better forecasts than transfer scaling. This suggests that the use of transfer scaling approach contributes little, but a closer look identifies the following implications (without statistical test). When the sample sizes from the older and the recent time points are very large and very small, respectively, the transfer scaling produces better forecasts. In addition, the distribution of L1 has less variance than that of L2, so the results are more robust. Furthermore, transfer scaling is less problematic regarding the poor estimates caused by the smaller number of observations from the recent time point.

[References]

Badoe, D.A., and E.J. Miller. Comparison of Alternative Methods for Updating Disaggregate Logit Mode Choice Models. Transportation Research Record, No. 1493, 1995, pp. 90-100.
Karasmaa, N., and M. Pursula. Empirical Studies of Transferability of Helsinki Metropolitan Area Travel Forecasting Models. Transportation Research Record, No. 1607, 1997, pp. 38-44.