A Balls-and Bins Model of Trade

Increase in trade caused by establishment of new trade routes or categories is called extensive margin, whereas increasing trade volume of existing trade routes/categories is called intensive margin. To better study the extensive margin of international trades, Roc Armenter and Miklos Koren proposed a new statistical model in the article *A Balls-and-Bins Model of Trade* to account for the sparsity of data (Armenter, 2014).

Formally, let $n \in \mathbb{N}$ be the number of balls (observations). Let $K \in \mathbb{N}$ denote the number of bins (categories), each of them indexed by subscript $i \in \{1,2,\dots,K\}$. The probability that any given ball lands in bin i is given by the bin size s_i, with $0 < s_i < 1$ and $\sum_{i=1}^{K} s_i = 1$. Then, for given bin-size distribution $\{s_i\}$, the joint probability of a number of balls $\{x_i\}$ is given by the multinomial distribution:

$$\Pr(x_1, \dots, x_K) = \frac{n!}{x_1!\, x_2! \dots x_K!} {s_1}^{x_1} \dots {s_K}^{x_K}$$

If we denote the number of nonempty bins by k, then we have:

$$k = \sum_{i=1}^{K} d_i$$

Where $d_i = 1$ if bin i is nonempty, and $d_i = 0$ otherwise. And the expectation of k is:

$$E(k\,|n) = \sum_{i=1}^{K} [1 - (1 - s_i)^n]$$

In this model, the bin-size distribution $\{s_i\}$, the number of observations n, and the number of categories K are all exogenous variables, while the number of nonempty categories k, and the indicator variables $\{d_i\}$ are endogenous variables.

The model presented above is a linear, static, stochastic model. The model is linear because it does not have nonlinearity, which is commonly associated with irreversibility or chaos. Moreover, the model is a static model because it does not account for time-dependent changes in the system. Finally, the model above is a stochastic model, because the outcome k is not determined not only by inputs, but also by random processes.

One factor that the model has not taken into consideration is the time-dependent variation of bin sizes. As a matter of fact, because each shipment in trade occurs at a different point of time (as do the balls fall in to the bins), the distribution of bin sizes are subject to change as matter of time in real life. While introducing time-dependent variation of bin sizes may help better increase the model accuracy, it may also lead to an unnecessarily complex model.

References Cited:

1. Armenter, Roc, and Miklós Koren. "A Balls-and-Bins Model of Trade." *The American Economic Review*, vol. 104, no. 7, 2014, pp. 2127–2151.