Identification and Positive Definiteness in the Joint Model

Supplemental note to
"A Joint Model of Residential Relocation Choice and Underlying Causal Factors"
by Katherine Kortum, Rajesh Paleti, Chandra Bhat, and Ram Pendyala

<u>Identification</u>
All notations in this note are the same as in the main paper. As only utility differences matter in a model with nominal dependent variables, the estimation process proceeds by considering the parameters only in the covariance matrix $\breve{\Sigma}^*_g$ of utility differences taken with respect to the first alternative for each of the two nominal variables[1]. Then, the variance term at the top left diagonal of $\breve{\Sigma}^*_g$ (g=1,2) is set to one to account for scale invariance. Next, consider the overall covariance matrix of utility differences (taken with respect to the first alternative for each nominal variable). This complete covariance matrix of utility differences of dimension $\tilde{G}*\tilde{G}$ is given by:

$$\breve{\Sigma}^* = \begin{bmatrix} \breve{\Sigma}^*_1 & \breve{\Sigma}^*_{12} \\ \breve{\Sigma}^*_{21} & \breve{\Sigma}^*_2 \end{bmatrix} \tag{7}$$

All elements of the matrix $\breve{\Sigma}^*$ are identifiable, and are the ones estimated. In the general case, this allows the estimation of $\sum_{g=1}^{2}\left(\dfrac{I_g*(I_g-1)}{2}-1\right)$ terms from $\breve{\Sigma}^*_1$ and $\breve{\Sigma}^*_2$, and $(I_1-1)\times(I_2-1)$ covariance terms in the off-diagonal matrices of the $\breve{\Sigma}^*$ matrix characterizing the dependence between the latent utility differentials (with respect to the first alternative) of the two nominal variables. The general covariance matrix Ω for the original (I_1+I_2)-error term vector $\varepsilon=(\varepsilon_1',\varepsilon_2')'$ can then be obtained as $\Omega=D\breve{\Sigma}^*D'$, where the matrix D is given by:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{11*9}$$

---

[1] Note that the matrix $\breve{\Sigma}^*_g$ is different from the matrix Σ^*_g, which corresponds to the covariance of utility differences taken with respect to the chosen alternative for the individual.

The covariance matrix $\mathbf{\Sigma}^*$ that is needed for estimation (and includes elements with respect to each household's chosen alternative for each nominal variable) can be obtained as $\mathbf{\Sigma}^* = \mathbf{M\Omega M}'$, where M is constructed as follows. The first $(I_1 - 1)$ rows and I_1 columns correspond to the first nominal variable. Insert an identity matrix of size $(I_1 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. The rest of the columns for the first $(I_1 - 1)$ rows and the rest of the rows for the first I_1 columns take a value of zero. Next, rows (I_1) through $(I_1 + I_2 - 2)$ and columns $(I_1 + 1)$ through $(I_1 + I_2)$ correspond to the second nominal variable. Again, position an identity matrix of size $(I_2 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative.

Positive Definiteness

The positive definiteness of the covariance matrix $\mathbf{\Sigma}^*$ (which is constructed multiple times during the optimization routine of the estimation process) can be ensured by making certain that the covariance matrix $\breve{\mathbf{\Sigma}}^*$ of utility differences taken with respect to the first alternative of each of the nominal variables is positive definite. This is done by using the Cholesky matrix of $\breve{\mathbf{\Sigma}}^*$ as the matrix of parameters to be estimated. Also, the Cholesky elements must be parameterized appropriately to ensure that the top diagonal element of each of $\breve{\mathbf{\Sigma}}_1^*$ and $\breve{\mathbf{\Sigma}}_2^*$ is normalized to one.

In the current empirical context, since there are only two alternatives in the first nominal variable, $\breve{\mathbf{\Sigma}}_1^* = 1$. The off-diagonal elements in $\breve{\mathbf{\Sigma}}_2^*$ account for the possible correlation in the utilities associated with different alternatives in the choice set corresponding to the drivers of residential choice. For example, a household which values aesthetics may have a high preference for scenic locations. Also, given the household's aesthetic sensitivity, the household may also value attributes of home quality and the neighborhood quality. This leads to a correlation in the utilities of the "scenic", "home quality", and "neighborhood" alternatives because the aesthetic preferences of household are usually not observed and thus not accounted for in the observed portion of the utilities. The elements in $\breve{\mathbf{\Sigma}}_{12}^*$, on the other hand, account for possible sample-selection effects, as discussed earlier.