Approaches to Model Transferability and Updating: The Combined Transfer Estimator

MOSHE BEN-AKIVA AND DENIS BOLDUC

The idea of model transferability is to use previously estimated model parameters from a different area for model estimation. The combined transfer estimator is based on the mean squares error criterion and extends the Bayesian procedure to explicitly account for the presence of a transfer bias. The suggested estimator is easy to apply because it is expressed as a linear combination of the direct estimation results and the previously estimated parameters. The combined estimator is shown to have superior accuracy in a mean square error sense to a direct (unbiased nontransfer) estimator whenever the transfer bias is relatively small. Numerical examples of the transfer region—where the combined estimator is superior to the direct estimator—are provided.

Model transferability is a practical approach to the problem of estimating a model for a study in an area for which the size of the available sample is small [for detailed discussions of transferability methods, see Ben-Akiva *(1)* and Koppelman and Wilmot *(2)*]. The model transfer approach is based on the idea that estimated model parameters from a previous study in a different area may provide useful information for estimating the parameters for the same model in a new area, even when the true values of the parameters are not expected to be equal. In the present notation, β_1 and β_2 denote the true $(K \times 1)$ parameter vectors of Areas 1 and 2 (the new area), respectively. The difference $\Delta = \beta_1 - \beta_2$ is called the transfer bias. In model transferability, one attempts to use the estimated parameters from Area 1, denoted by b_1, to improve the accuracy of the estimation of β_2. The difficulty occurs because b_1 is an estimator of β_1 and not of β_2.

If the transfer bias is negligible (as indicated, for example, by a Chow test of the null hypothesis that the vectors, or specific subvectors, of the model parameters for the two areas are equal), the two data sets can be pooled, and identical parametric values can be estimated for the two areas. Alternatively, and particularly if the original data for Area 1 are not available, a Bayesian updating procedure can be used *(3)*. However, these pooled and Bayesian estimators are not appropriate for situations in which coefficients for the transfer bias cannot be assumed to be negligible. The transfer scaling approach previously applied in these situations is described in the following section. It takes the estimator b_1 and attempts a correction of the transfer bias by using the data from Area 2.

In this paper, a new model transfer estimator, the combined transfer estimator, is developed. It is stated as a weighted average of the direct estimators b_1 and b_2. The term direct estimator is used in this paper to mean an unbiased estimation procedure that would be performed if no transfer was attempted. The weights are assigned in such a way that for each value of the transfer bias, the mean square error (MSE) of the combined estimator is minimized. This approach is expected to yield better estimates of β_2 when the transfer bias is small.

M. Ben-Akiva, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Mass. 02139. D. Bolduc, Département d'économique, Université Laval, Québec G1K 7P4, Canada.

THE TRANSFER SCALING APPROACH

A relationship between the true values of the parameters in the two areas is called a transfer bias model. Consider the following:

$$\beta_2 = \text{diag}(\mu)\beta_1 = \text{diag}(\beta_1)\mu \tag{1}$$

where μ is a $(K \times 1)$ vector of unknown bias scale parameters and $\text{diag}(\mu)$ and $\text{diag}(\beta_1)$ are $(K \times K)$ diagonal matrices in which the *kk*th elements are μ_k and β_{1k}, respectively. For simplicity, the matrix $\text{diag}(\mu)$ will be denoted by M. The relation $\beta_2 = \text{M}\beta_1$ is such that the transfer bias, $(\beta_1 - \beta_2) = (\text{I} - \text{M})\beta_1$, is nonzero unless $\mu_k = 1, \forall k = 1, \ldots, K$. Denote the number of distinct parameters in μ by M. In general, $M \leq K$, and the usefulness of the transfer scaling approach, described in the following, is for cases where $M < K$. Gunn et al. *(4)* thoroughly tested the transfer scaling approach by classifying the independent variables of a travel demand model into groups with similar transfer bias properties.

Equation 1 can also be expressed as

$$\begin{aligned}\beta_2 &= \text{M}\beta_1 = \text{M}b_1 + \text{M}(\beta_1 - b_1)\\ &= \text{diag}(b_1)\mu + M(\beta_1 - b_1)\end{aligned} \tag{2}$$

If b_1 is an unbiased estimator of β_1, the vector $\text{M}(\beta_1 - b_1)$ is a simple transformation of the sampling error in Area 1 estimates and has an expected value of zero. In a transfer scaled model, μ denotes the vector of parameters that are estimated from Area 2 data. If m is called the vector of estimates for μ, the transfer scaling estimator of β_2 is computed as $\text{diag}(b_1)m$. In all previous applications of the transfer scaling approach, the term $\text{M}(\beta_1 - b_1)$ has been ignored. In the estimation with Area 2 data, it represents measurement errors in the independent variables. This term plays a critical role because unless it is negligible, the transfer scaling estimator is inefficient and potentially biased because it is correlated with the independent variables in the model estimated with Area 2 data.

The transfer scaling approach can be implemented at different levels of detail with M (the number of bias scale parameters) ranging from none (i.e., assuming $\beta_1 - \beta_2 = 0$ by imposing $\mu_1 = \mu_2 = \ldots = \mu_K = 1$) to K (i.e., one bias scale parameter per model parameter). For $M = K$, the transfer scale estimator $\text{diag}(b_1)m$ is identical to Area 2 direct estimator b_2.

Thus transfer scaling is a useful approach when the data available for the new area permit estimation of only a small number of new parameters and an accurate estimator, that is, one with small $(b_1 - \beta_1)$, is available from another area. The latter requirement is needed to justify the assumption that the term $M(\beta_1 - b_1)$ is negligible.

In a related paper, Ben-Akiva and Bolduc *(5)* also develop a second model transferability approach—a mixed estimation procedure that jointly estimates the new area model and a transfer bias model. This mixed estimation may be viewed as an extension of the transfer scaling approach that overcomes the deficiencies due to a nonnegligible $M(\beta_1 - b_1)$ value.

MINIMUM MSE ESTIMATOR

The objective of model transfer is to improve the estimation of β_2 by combining the sample information from Area 2 with knowledge of b_1. The transfer scaling approach achieves this objective by postulating a specific transfer bias model and estimating the parameters of the transfer bias model from Area 2 data. The Area 2 data are used only to correct the transfer bias, and differences in sampling errors between the two data sets are not explicitly recognized. By using the two direct estimates b_1 and b_2 directly, an estimator is developed that treates the trade-off between sampling errors and transfer biases explicitly.

The Problem

Given the direct estimators b_1 and b_2, find a combined estimator defined by the function

$$\ddot{b}_2 = h(b_2, b_1)$$

that in some sense is a better estimator of β_2 than the direct estimator b_2.

We use the MSE criterion, which implies, for example, that for a variance reduction one is willing to allow a bias. A brief description of optimal MSE estimation with a single parameter follows. A more extended treatment can be found in the work of Judge et al. *(6)*.

The Minimum MSE Approach

Use the MSE criterion to find whether

$$\text{MSE}(\ddot{b}_2) \leq \text{MSE}(b_2)$$

or

$$E(\ddot{b}_2 - \beta_2)^2 \leq E(b_2 - \beta_2)^2 \tag{3}$$

holds. An MSE is equal to the square of the bias plus the variance. Because b_2 is assumed to be unbiased, the criterion is reduced to

$$E(\ddot{b}_2 - \beta_2)^2 \leq \text{Var}\,(b_2) \tag{4}$$

or

$$\text{Var}\,(\ddot{b}_2) + B^2(\ddot{b}_2) \leq \text{Var}\,(b_2) \tag{5}$$

where the bias of $\ddot{b}_2$ is defined as $B(\ddot{b}_2) = E(\ddot{b}_2) - \beta_2$ and the variance definition is, for example, $\text{Var}\,[\ddot{b}_2 - E(\ddot{b}_2)]^2$. This inequality, demonstrated in Figure 1, reveals how the advantage gained from variance reduction [i.e., $\text{Var}(\ddot{b}_2) \leq \text{Var}(b_2)$] may be significantly reduced or even totally lost by the presence of a significant transfer bias.

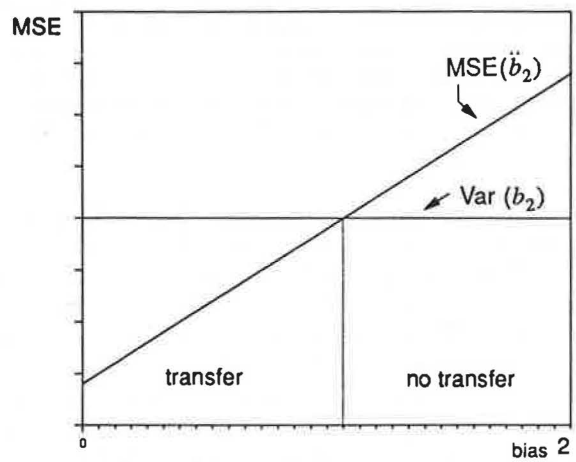

FIGURE 1 When to transfer as a function of the transfer bias.

For a model with K parameters, the combined transfer estimator developed later in this paper is based on a weighted average of the two direct estimators (i.e., h is assumed to be a linear function) and is expressed as follows:

$$\ddot{b}_2 = (I - A)b_2 + Ab_1$$

or

$$\ddot{b}_2 = b_2 + A(b_1 - b_2) \tag{6}$$

where A is a $(K \times K)$ matrix of weights. The matrix A is a general matrix for which element a_{ij} of matrix A gives the relative importance of b_{1j} in the estimation of $\ddot{b}_{2i}$.

THE COMBINED TRANSFER ESTIMATOR

The MSE optimal value for the weighting matrix A of the combined estimator can now be obtained. First, the one parameter case is developed in detail and then the derivation is extended to the multiparameter case.

Derivation of the Combined Transfer Estimator

For $K = 1$, the combined transfer estimator is expressed as a linear combination of the direct estimators with fixed weights, as follows:

$$\ddot{b}_2 = (\alpha_1 + \alpha_2)^{-1} \alpha_1 b_1 + (\alpha_1 + \alpha_2)^{-1} \alpha_2 b_2$$
$$= b_2 + \alpha(b_1 - b_2) \qquad (7)$$

where $\alpha_1, \alpha_2 \geq 0$ and $\alpha = (\alpha_1 + \alpha_2)^{-1}\alpha_1$. This is a non-Bayesian estimator that combines the information from the two samples. In a Bayesian setting the random vector b_1 would be replaced by the fixed mean of the prior distribution of β_2.

The expected value of $\ddot{b}_2$ is

$$E(\ddot{b}_2) = \beta_2 + \alpha(\beta_1 - \beta_2) \qquad (8)$$

The MSE optimal value for α is obtained by minimizing the MSE as a function of α:

$$\underset{\alpha}{\text{Minimize}}\ \text{MSE} = [\text{Var}(\ddot{b}_2) + B^2(\ddot{b}_2)]$$
$$= \alpha^2\sigma_1^2 + (1 - \alpha)^2\sigma_2^2 + \alpha^2(\beta_1 - \beta_2)^2 \quad (9)$$

where

$$\sigma_i^2 = \text{Var}(b_i) \qquad i = 1, 2$$

The first-order condition is

$$\frac{\partial \text{MSE}}{\partial \alpha} = 2\alpha\sigma_1^2 - 2(1 - \alpha)\sigma_2^2 + 2\alpha\Delta^2 = 0$$

which implies

$$\alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2 + \Delta^2} = \frac{\sigma_2^2}{D} \qquad (10)$$

where $\Delta = \beta_1 - \beta_2$ and $D = \sigma_1^2 + \sigma_2^2 + \Delta^2$. Because α is a function of Δ (which is an unknown quantity), in an empirical application Δ will have to be replaced by an observed quantity. Use of $d = b_1 - b_2$ is suggested. Another possibility is to apply the transfer scaling approach to estimate the transfer bias. Therefore, α is random in practice. The implications of this fact will be analyzed later on.

At this value of α, the optimal combined estimator is expressed as

$$\ddot{b}_2 = b_2 + \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2 + \Delta^2}(b_1 - b_2)$$

Multiplying α by

$$\frac{(\sigma_1^2 + \Delta^2)^{-1}\sigma_2^{-2}}{(\sigma_1^2 + \Delta^2)^{-1}\sigma_2^{-2}}$$

implies

$$\alpha = \frac{(\sigma_1^2 + \Delta^2)^{-1}}{(\sigma_1^2 + \Delta^2)^{-1} + \sigma_2^{-2}}$$

For $\Delta = 0$, the weights are obtained in the Bayesian updating formula, which is MSE optimal if the transfer bias is zero.

Properties of the Combined Transfer Estimator

The optimal combined estimator $\ddot{b}_2$ just derived can be compared with b_2, the direct estimator. Substitution of α from Equation 10 in the objective function of Equation 9 yields

$$\text{MSE}(\ddot{b}_2) = \alpha\sigma_1^2\frac{\sigma_2^2}{D} + (1 - \alpha)\left[\frac{\sigma_1^2 + \Delta^2}{D}\right]\sigma_2^2 + \alpha\frac{\sigma_2^2}{D}\Delta^2$$
$$= \alpha\left[\frac{\sigma_1^2 + \Delta^2}{D}\right]\sigma_2^2 + (1 - \alpha)\left[\frac{\sigma_1^2 + \Delta^2}{D}\right]\sigma_2^2$$
$$= \left[\frac{\sigma_1^2 + \Delta^2}{D}\right]\sigma_2^2$$
$$= (1 - \alpha)\sigma_2^2 \qquad (11)$$

Because $0 \leq \alpha \leq 1$,

$$\text{MSE}(\ddot{b}_2) \leq \text{Var}(b_2) \qquad \forall\Delta^2 \geq 0 \qquad (12)$$

Thus the optimal combined estimator always stays in the transfer region. It is always better because as the bias increases α decreases, and for $|\beta_1 - \beta_2| \rightarrow \infty$, $\alpha \rightarrow 0$, and $\ddot{b}_2 \rightarrow b_2$. The pattern of $\text{MSE}(\ddot{b}_2)$ as a function of the transfer bais is investigated next. The first and second derivatives with respect to Δ^2 are as follows:

$$\frac{\partial \text{MSE}(\ddot{b}_2)}{\partial \Delta^2} = \frac{\partial \text{MSE}(\ddot{b}_2)}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \Delta^2} = \alpha^2 \geq 0$$

$$\frac{\partial^2 \text{MSE}(\ddot{b}_2)}{(\partial \Delta^2)^2} = \frac{\partial \alpha^2}{\partial \Delta^2} = -2\frac{\alpha^2}{D} \leq 0$$

This is because

$$\frac{\partial \alpha}{\partial \Delta^2} = -\frac{\sigma_2^2}{D^2} = -\frac{\alpha}{D}$$

At $\Delta = 0$, $\text{MSE}(\ddot{b}_2) = \sigma_2^2[\sigma_1^2/(\sigma_1^2 + \sigma_2^2)]$. The pattern of $\text{MSE}(\ddot{b}_2)$ as a function of Δ^2 is shown in Figure 2.

The results of the analysis have been obtained under the hypothesis that α is a known fixed constant. When Δ is replaced by an estimate d (i.e., a random variable), it becomes difficult to evaluate $E(\ddot{b}_2)$ and Var $(\ddot{b}_2)$. In what follows, these moments are approximated by a Taylor's series expansion. It will be shown that in the case of an estimated α it is not impossible for the combined estimator to be inferior to the direct estimator. In this analysis, σ_1^2 and σ_2^2 are assumed to be known, and therefore the randomness in α arises only from the substitution of d for Δ.

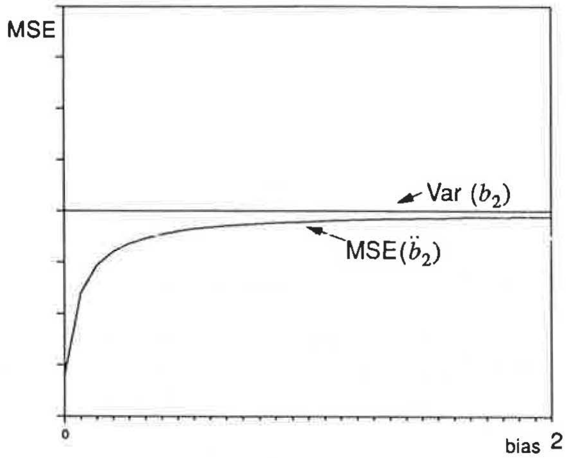

FIGURE 2 The combined estimator is always superior.

Recall the combined estimator in Equation 10, and replace Δ with its estimator d, as follows:

$$\tilde{b}_2 = b_2 + a(b_1 - b_2) \quad (13)$$

with

$$a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2 + d^2} = \frac{\sigma_2^2}{\hat{D}}$$

where

$$\hat{D} = \sigma_1^2 + \sigma_2^2 + d^2$$

The $\tilde{b}_2$ is a nonlinear function of the random variables b_1, b_2, and d. The moments of $\tilde{b}_2$ are approximated by a Taylor's series expansion around the true values of β_1 and β_2 (as well as $\Delta = \beta_1 - \beta_2$), as follows:

$$E(\tilde{b}_2) = \beta_2 + \alpha\Delta \quad (14)$$

and

$$\mathrm{Var}(\tilde{b}_2) = J'\Sigma J \quad (15)$$

where

$$J' = [\partial\tilde{b}_2/\partial b_1, \partial\tilde{b}_2/\partial b_2]_{\beta_1,\beta_2}$$

and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

The calculations are performed under the assumption

$$d = b_1 - b_2 \quad (16)$$

The partial derivatives of $\tilde{b}_2$ in J are derived by Ben-Akiva and Bolduc *(7)* and lead to

$$J' = [\alpha - 2\Delta^2\alpha/D, 1 - \alpha + 2\Delta^2\alpha/D] \quad (17)$$

Substitution of Equation 17 into Equation 15 yields

$$\mathrm{Var}(\tilde{b}_2) = \sigma_1^2\alpha^2(1 - 2\Delta^2/D)^2 + \sigma_2^2(1 - \alpha)^2[1 + 2\Delta^2\alpha/D(1 - \alpha)]^2$$

The MSE is

$$\mathrm{MSE}(\tilde{b}_2) = \alpha^2\Delta^2 + \mathrm{Var}(\tilde{b}_2)$$

At the two limits $\Delta^2 = 0$ and $\Delta^2 \to \infty$, this MSE expression coincides with the one obtained before for deterministic α. However, depending on the values of the parameters, it is now possible that for a finite value of Δ^2 the MSE of the combined estimator, MSE($\tilde{b}_2$), will exceed the variance of the Area 2 direct estimator, Var(b_2). This is demonstrated in the numerical example in Figure 3 for $\sigma_1^2 = 1$ and $\sigma_2^2 = 4$.

The combined estimator presents an improvement over a direct estimator only within a limited transfer region in which the transfer bias is relatively small. If the transfer bias is greater than some critical value, the combined estimator is, in fact, inferior to the direct estimator. Thus in practical applications the transfer region for the optimal combined estimator is not global. Clearly, any application of the notion of transferability is based on the prior assumption that the transfer bias is relatively small, or in other words, that there are a priori expectations that the model parameters are similar between the two areas.

The sensitivity of the transfer region to the values of σ_1^2 and σ_2^2 is demonstrated in Figures 4 and 5, respectively. In general, it is shown that increasing σ_1^2 or σ_2^2 leads to a larger transfer region. Figure 4 shows that as σ_1^2 gets larger, the gain in accuracy from the transfer estimator is reduced in situations with small transfer bias. In general, as σ_1^2 increases, the value of α decreases, and the combined estimator approaches the direct estimator. Thus for $\sigma_1^2 \to \infty$ (inaccurate information from Sample 1) the MSE($\tilde{b}_2$) curve approaches the horizontal line of σ_2^2. The dramatic effects of σ_2^2 on the transfer region and the accuracy of the combined estimator are demonstrated in Figure 5. The size of the transfer region appears to be more sensitive to σ_2^2 than to σ_1^2, and at the limit for $\sigma_2^2 \to \infty$ the transfer region is obviously global.

Monte Carlo experiments were performed to evaluate the accuracy of the expression for MSE($\tilde{b}_2$) that was developed under the assumption of known σ_1^2 and σ_2^2 and a first-order Taylor's series approximation. The experiments, described by Ben-Akiva and Bolduc *(7)*, compare the true MSE and the MSE curve computed by using the previous approximations. The results clearly show that the approximate MSE curve significantly underestimates the point at which the true MSE curves intersect with σ_2^2. In other words, the approximate MSE curve provides a conservative estimate of the transfer region. The Monte Carlo results show that the critical value may be as

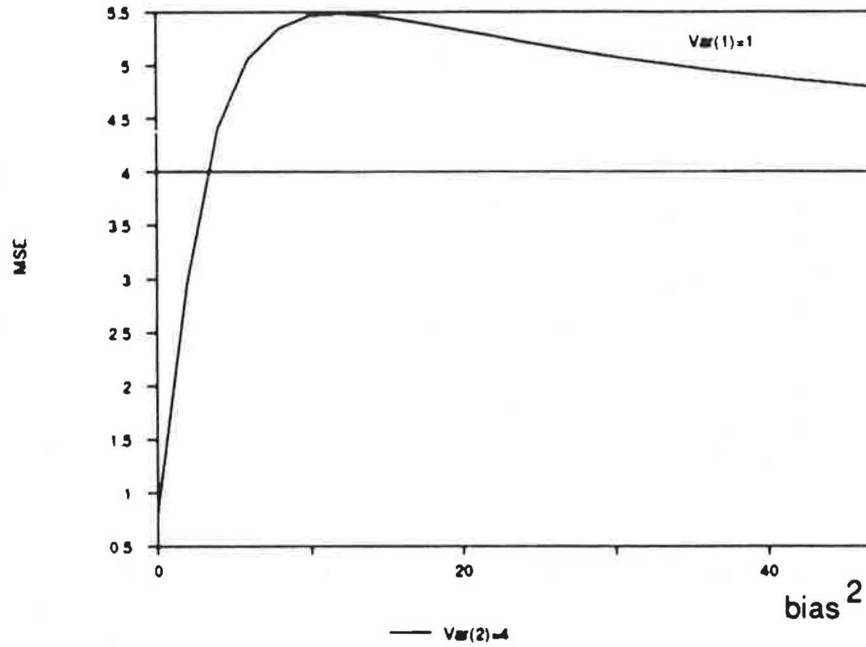

FIGURE 3 The transfer region of the combined estimator.

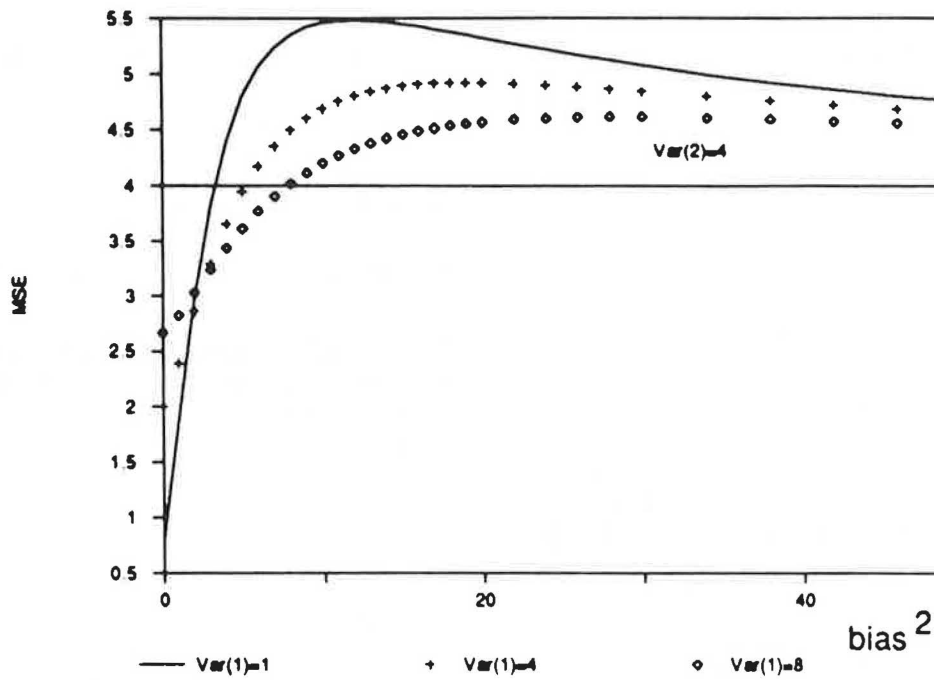

FIGURE 4 The effect on the transfer region of Var *(1)*.

high as 3 or 4 times greater than the one obtained from the approximate analysis. This result should be taken into account in empirical applications.

THE MULTIVARIATE EXTENSION OF THE COMBINED ESTIMATOR

Here, the derivation of the combined transfer estimator is extended to the multivariate case. The combined estimator is defined with nonsingular fixed-weight matrices A_1 and A_2 as follows:

$$\ddot{b}_2 = (A_1 + A_2)^{-1}A_1b_1 + (A_1 + A_2)^{-1}A_2b_2 \quad (18)$$

Let

$$A = (A_1 + A_2)^{-1}A_1$$

and note that

$$(A_1 + A_2)^{-1}A_2 = I - A$$

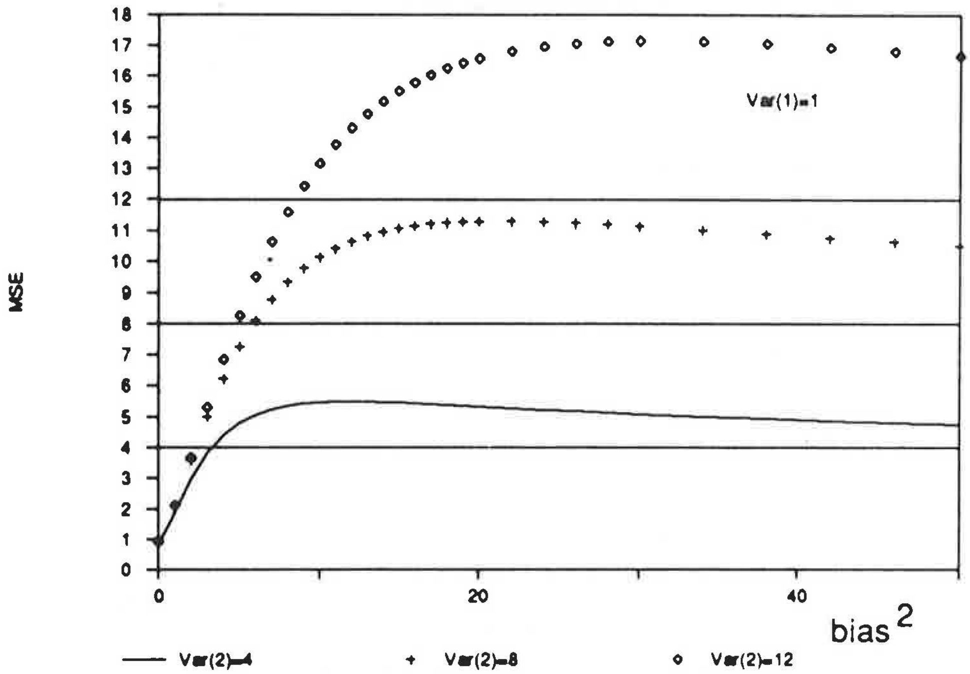

FIGURE 5 The effect on the transfer region of Var *(2)*.

Rewrite the combined estimator (Equation 18) as

$$\ddot{b}_2 = (I - A)b_2 + Ab_1 = b_2 + A(b_1 - b_2)$$

and obtain the following expectation of $\ddot{b}_2$:

$$E(\ddot{b}_2) = \beta_2 + A(\beta_1 - \beta_2) \quad (19)$$

Let $\Delta = \beta_1 - \beta_2$ and express the bias of $\ddot{b}_2$ by

$$B(\ddot{b}_2) = E(\ddot{b}_2) - \beta_2 = A\Delta$$

Denote the covariance matrices of the direct estimators by $\text{Var}(b_i) = \Sigma_i$, $i = 1, 2$. The two samples are independent, and as a consequence, b_1 and b_2 are independently distributed. Under this hypothesis, the covariance matrix of the combined estimator is given by

$$\begin{aligned}\text{Var}(\ddot{b}_2) &= (I - A)\Sigma_2(I - A)' + A\Sigma_1A' \\ &= \Sigma_2 - \Sigma_2A' - A\Sigma_2 + A\Sigma_2A' + A\Sigma_1A'\end{aligned}$$

The latter results in the following MSE expression for $\ddot{b}_2$:

$$\begin{aligned}\text{MSE}(\ddot{b}_2) = \Sigma_2 - \Sigma_2A' - A\Sigma_2 \\ + A(\Sigma_1 + \Delta\Delta' + \Sigma_2)A'\end{aligned} \quad (20)$$

Define the optimal weighted average estimator as the matrix A that minimizes the trace of the $\text{MSE}(\ddot{b}_2)$ matrix, as follows:

$$\begin{aligned}\text{tr}\,[\text{MSE}(\ddot{b}_2)] = \text{tr}\,[\Sigma_2] - 2\,\text{tr}\,[\Sigma_2A'] \\ + \text{tr}\,[A(\Sigma_1 + \Delta\Delta' + \Sigma_2)A']\end{aligned}$$

The optimal value of matrix A is given by

$$A = \Sigma_2(\Sigma_1 + \Delta\Delta' + \Sigma_2)^{-1} \quad (21)$$

which can also be written as

$$A = [(\Sigma_1 + \Delta\Delta')^{-1} + \Sigma_2^{-1}]^{-1}\,(\Sigma_1 + \Delta\Delta')^{-1}$$

For a detailed derivation, see Ben-Akiva and Bolduc *(7)*. Note that in the scalar case (e.g., $K = 1$), the matrix A reduces to the value of α derived earlier.

Equation 21 implies that the optimal weight matrices can be taken to be

$$A_1 = (\Sigma_1 + \Delta\Delta')^{-1}$$

$$A_2 = \Sigma_2^{-1}$$

Substitution of these matrices in the estimator (Equation 18) yields

$$\begin{aligned}\ddot{b}_2 = [(\Sigma + \Delta\Delta')^{-1} + \Sigma_2^{-1}]^{-1}\,[(\Sigma_1 \\ + \Delta\Delta')^{-1}b_1 + \Sigma_2^{-1}b_2]\end{aligned} \quad (22)$$

As in the scalar case for $\Delta = 0$, this estimator reduces to the Bayesian updating formula.

The approach used in the single parameter case is now used to derive an expression for $\text{MSE}(\tilde{b}_2)$ when the matrix A is replaced by an estimate $\hat{A}$. As in the one-parameter case, assume that Σ_1 and Σ_2 are known and that the randomness in $\hat{A}$ arises from the use of d, which is an estimate of Δ. As suggested earlier, the most straightforward estimate of Δ is the difference between the direct estimates: $d = b_1 - b_2$.

Recall matrix A in Equation 21 and replace Δ with its estimate d, as follows:

$$\tilde{b}_2 = b_2 + \hat{A}(b_1 - b_2)$$

with

$$\hat{A} = \Sigma_2(\Sigma_1 + dd' + \Sigma_2)^{-1} = \Sigma_2\hat{D}^{-1}$$

where

$$\hat{D} = \Sigma_1 + dd' + \Sigma_2$$

The Taylor's series approximation yields the following:

$$E(\tilde{b}_2) = \beta_2 + A\Delta$$

$$\mathrm{Var}(\tilde{b}_2) = J'\Sigma J$$

where

$$J' = [\partial\tilde{b}_2/\partial b_1', \partial\tilde{b}_2/\partial b_2']_{\beta_1,\beta_2}$$

$$\Sigma = \begin{pmatrix}\Sigma_1 & 0 \\ 0 & \Sigma_2\end{pmatrix}$$

The partial derivatives of $\tilde{b}_2$ are

$$\partial\tilde{b}_2/\partial b_1' = \hat{A} - d'\hat{D}^{-1}d \otimes \hat{A} - d'\hat{D}^{-1} \otimes \hat{A}d$$

and

$$\partial\tilde{b}_2/\partial b_2' = I - \hat{A} + d'\hat{D}^{-1}d \otimes \hat{A} + d'\hat{D}^{-1} \otimes \hat{A}d$$

where $\otimes$ denotes a Kronecker product [for details, see Ben-Akiva and Bolduc *(7)*]. These expressions can be used to derive the multivariate transfer regions that are useful in situations with significant off-diagonal elements in Σ_1 and Σ_2.

CONCLUSION

A new approach to model transferability based on the MSE criterion was developed. The combined transfer estimator was derived, and it was shown that for sufficiently small transfer bias it dominates the direct estimator of the model in a new area. The combined estimator may be viewed as an extension of the Bayesian updating procedure that explicitly accounts for the possible presence of a transfer bias. The computational requirements of the combined transfer estimator are the same as those of the transfer scaling or the Bayesian updating procedures.

A linear approximation was employed to analyze the properties of the combined transfer estimator. However, results of Monte Carlo experiments have shown that the linear approximation underestimates the improvement of the combined estimator over the direct estimation. To overcome this problem, it will be necessary in further research to develop exact distributional results for the combined estimator.

Another approach that overcomes the statistical deficiencies of the transfer scaling approach but is computationally more demanding is to view transferability as a mixed estimation problem. In a related paper, a mixed estimator is proposed that jointly estimates the new area model and the transfer bias model used in the transfer scaling approach *(5)*.

ACKNOWLEDGMENT

The authors have benefited from discussions with Bruno Boccara, Andrew Daly, Hugh Gunn, Dan McFadden, and Rohit Ramaswamy.

REFERENCES

1. M. Ben-Akiva. Issues in Transferring and Updating Travel Behavior Models. In *New Horizons in Travel Behavior Research.* (P. R. Stopher, A. G. Meyburg, and W. Brog, eds.), Lexington Books, Lexington, Mass., 1980.
2. F. S. Koppelman and C. G. Wilmot. Transferability Analysis of Disaggregate Choice Models. In *Transportation Research Record 895*, TRB, National Research Council, Washington, D.C., 1983, pp. 18–24.
3. T. Atherton and M. Ben-Akiva. Transferability and Updating of Disaggregate Travel Demand Models. In *Transportation Research Record 610*, TRB, National Research Council, Washington, D.C., 1976, pp. 12–18.
4. H. F. Gunn, M. Ben-Akiva, and M. A. Bradley. Tests of the Scaling Approach to Transferring Disaggregate Travel Demand Models. In *Transportation Research Record 1037*, TRB, National Research Council, Washington, D.C., 1985, pp. 21–30.
5. M. Ben-Akiva and D. Bolduc. *Approaches to Model Transferability: Combined and Mixed Estimators.* Working paper, M.I.T., Cambridge, Mass., 1985.
6. G. G. Judge, W. E. Griffiths, R. C. Hill, and T. C. Lee. *The Theory and Practice of Econometrics.* John Wiley and Sons, New York, 1980.
7. M. Ben-Akiva and D. Bolduc. *The Combined Estimator Approach to Model Transferability and Updating.* Working paper, M.I.T., Cambridge, Mass., 1987.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.