"RECURSIVE LEAST-SQUARES APPROACH TO DATA TRANSFERABILITY: EXPOSITION AND NUMERICAL RESULTS"

by

Lydia J Price*
and
Wilfried R Vanhonacker**

N° 92/38/MKT

* Assistant Professor of Marketing at INSEAD,
Boulevard de Constance, Fontainebleau 77305 Cedex, France.

** Professor of Marketing at INSEAD,
Boulevard de Constance, Fontainebleau 77305 Cedex, France.

Printed at INSEAD,
Fontainebleau, France

RECURSIVE LEAST-SQUARES APPROACH TO DATA TRANSFERABILITY: EXPOSITION AND NUMERICAL RESULTS

Lydia J. PRICE

and

Wilfried R. VANHONACKER*

Revised May 1992

* Assistant Professor and Professor of Marketing, INSEAD, Fontainebleau, France.
Funding was provided by the R&D Department of INSEAD. This research benefited from critical comments by colleagues at Columbia University and INSEAD.

Recursive Least-Squares Approach to Data Transferability:

Exposition and Numerical Results

ABSTRACT

Data transferability refers to the transfer of information from a set of independently estimable models to a new but structurally equivalent model for which data on some predictors are missing. Using a random-coefficient regression framework, this paper discusses a recursive least-squares approach to execute the transfer and estimate all parameters of the new model. Numerical results document the method's sensitivity to practically relevant dimensions and, as such, establish its general applicability.

Key words: TRANSFERABILITY; RANDOM COEFFICIENT REGRESSION; RECURSIVE LEAST-SQUARES.

I INTRODUCTION

It is not uncommon to encounter a situation in which a proposed econometric model cannot be estimated because there are no observations on one or more of the model's predictor variables. In such a situation, however, it may be the case that parameter estimates are available from data sets pertaining to similar problems. The objective of data transferability is to supplement the incomplete data matrix for the current problem with the estimates obtained from similar problems, and thereby derive a set of estimates for *all* parameters in the econometric model, including those associated with unobservable predictors.

The problem described is one of historical analogy, i.e., the decision maker's task is to utilize information from historical cases to try to draw inferences about the current situation. As an example, consider the case of a drug that is about to come off patent. Given the likely prospect of substantial generic substitution upon patent expiration (Corstjens, 1991), the manager may try to anticipate the magnitude of the substitution effect by looking at other, similar drugs whose patents have already expired. The only knowledge that exists for the drug whose patent has not yet expired is the response history under patent protection. In a subjective and ad hoc sense, the manager may try to transfer information from the historically analgous situations into the latest drug's anticipated post-patent situation. Data transferability, however, offers a more formal methodology for doing this.

The data transferability problem has been studied by Aigner and Leamer (1984). They provide a concise representation of the problem, and offer an empirical Bayes approach as a solution. They also discuss at length a tracking measure which captures the amount of transfer that has occurred. Within a similar framework, Vanhonacker and Price (1992) discuss a recursive least-squares approach to the problem, which they then compare to the empirical Bayes approach of Aigner and Leamer (1984). As the focus of the

Vanhonacker and Price (1992) paper is on application of the transferability approach to a forecasting problem, there is little discussion of the behavior of the recursive least squares estimator.

The objective of this paper is to evaluate the recursive least-squares approach to data transferability. An extensive simulation study provides insight into the effects of 1) residual variance and 2) the number of predictors without observations on the efficiency of the transfer estimates and on the transfer measure as proposed by Aigner and Leamer (1984). The study also provides numerical insights about how the transferability model's covariance assumptions affect the transfer estimates and the transfer measure.

II - DATA TRANSFERABILITY

1. Problem Definition

In matrix algebra notation, we have a response model

$$y_o = X_o \beta_o + u_o \tag{1}$$

where y_o is (n x 1) containing dependent variable observations, X_o is (n x k) containing observations on k non-stochastic predictor variables, β_o is a (k x 1) parameter vector, and u_o is (n x 1) containing random disturbances. On some of the predictors in X_o we do not have any observations. X_o in (1) has zero entries in the columns corresponding to those predictors. As such, X_o is not a full-rank matrix and no independent least-squares estimate can be directly obtained for β_o.

We do, however, have access to m similar models for which the observations are known. Specifically, we have

$$y_i = X_i \beta_i + u_i \qquad \text{for } i = 1, 2, \ldots, m \tag{2}$$

with terms defined as in (1). For all i, matrix X_i is full rank (equal to k). For notational simplicity, we will assume that an equal number of observations are available for the variables in (2). The transferability methodology is not constrained to such a case but will require that there are enough observations for each of the m historical cases so that independent estimates for β_i ($i = 1, 2, ..., m$) can be obtained. Those estimates contain information with which we will update the partial information in X_o to estimate β_o in (1).

Of importance is to operationalise the similarity between (1) and the m models in (2). Furthermore, we will have to define a link which enables information to be transferred from (2) into (1). Similarity in (1) and (2) is realized when the models are structurally identical. Specifically, the (m + 1) models are assumed to have identical functional forms (e.g., linear, multiplicative, etc.) and to contain an identical set of predictor variables.

The link between the models is defined in a random-coefficients framework. Specifically, it is assumed that all parameter vectors (β_o and β_i for $i = 1, 2, ..., m$) come from the same multivariate density with mean vector $\bar{\beta}$ and variance-covariance matrix Σ, or

$$\beta_i \sim G(\bar{\beta}, \Sigma) \qquad \text{for } i = 0, 1, 2, ..., m. \qquad (3)$$

If the corresponding parameters are identical across situations, the diagonal elements (and, hence, the off-diagonal elements) of Σ will be zero and $\beta_i = \bar{\beta}$ for $i = 0, 1, 2, ..., m$. If they are different, the diagonal elements of Σ will be non-zero, representing the variance of the observed parameters around an overall mean contained in $\bar{\beta}$. In the latter instance, one would hope that the off-diagonal elements of Σ are non-zero too, as they would provide the link between the parameters within each situation. In absence of any such link, it would be impossible to make inferences about β_o based on the β_i's in (2).

The variance-covariance matrix Σ plays a crucial role in data transferability. It provides the link between the response coefficients of various instances. If matrix Σ is a null matrix, then the multivariate density in (3) collapses to the mean value $\bar{\beta}$ and all regression models in (1) and (2) have identical parameters. Accordingly, the most efficient estimate of β_o in (1) can be obtained from pooling the various time series. If, by contrast, the diagonal of Σ converges to an infinite level, then efficient estimates of the response coefficients can only be obtained from estimating each model in (1) and (2) separately. Note that the latter is true irrespective of the size of the covariances (i.e., the off-diagonal elements of Σ). Non-zero covariances indicate a potential for transfer, but infinitely large variances give rise to highly inefficient estimates. In other words, only noise would be transferred, which would not enhance the efficiency of the final estimates. Moreover, in the absence of information transfer, no estimate of β_o in (1) could be obtained as the singularity of matrix X_o prevents the independent estimation required for efficiency. Additional data would be needed (to make X_o essentially a full rank matrix) before β_o could be estimated efficiently.

2. The Recursive Least-Squares Estimator

Since historical data are available for all variables specified in (1), estimates can be obtained for the parameter vectors β_i for i = 1, 2, ..., m. Assuming that β_i is independent of u_i, and that the disturbances in u_i are independently and identically distributed with a mean zero and a variance-covariance matrix $\sigma_i^2 I$, we can rely on the extended Gauss-Markov theorem to obtain the OLS estimator of β_i, $\hat{\beta}_i = (X_i'X_i)^{-1} X_i'y_i$ which is the "best linear unbiased estimator" (BLUE) (see, e.g., Duncan and Horn 1972). These independently estimated vectors will now be combined with the partially available data contained in X_o to provide estimates for all parameters in β_o, including the ones associated with zero columns. A recursive estimator which treats the $\hat{\beta}_i$ estimates as priors accomplishes this task in a straight-forward Bayesian updating step.

To accomplish this, note that from (3), we have

$$\beta_i = \bar{\beta} + e_i \qquad \text{for } i = 1, 2, \ldots, m,$$

with $E(e_i) = 0$ and $E(e_i e_i') = \Sigma$. Accordingly,

$$\beta_i = \beta_o + V_{io} \qquad \text{for } i = 1, 2, \ldots, m$$

where $V_{io} = e_i - e_o$. Incorporating the $\hat{\beta}_i$ estimates discussed above, it can be shown that

$$\hat{\beta}_i = \beta_o + U_{io} \qquad \text{for } i = 1, 2, ..., m \qquad (4)$$

where $U_{io} = (X_i'X_i)^{-1}X_i'u_i + V_{io}$. Expression (4) states that the least-squares estimates $\hat{\beta}_i$ for i = 1, 2, ..., m can be interpreted as <u>prior estimates</u> of β_o, the parameter vector of model (1).

Since $E(u_i) = E(e_i) = 0$ for all i, $E(U_{io}) = 0$ for all i (assuming X_i for i = 1, 2, ..., m does not contain stochastic regressors). Assuming all random components are mutually independent, we have

$$E(U_{io}U_{io}') = \Sigma_i^* = 2\Sigma + \sigma_i^2(X_i'X_i)^{-1} \qquad \text{for } i = 1, 2, ..., m \qquad (5)$$

and

$$E(U_{io}U_{jo}') = \Sigma \qquad \text{for } i \neq j.$$

In other words, the $\hat{\beta}_i$'s contain first moment information on random vector β_o. The second moment information is contained in the variance-covariance matrix of U_{io} derived in (5).

All prior information contained in (4) can be summarized in a single equation as

$$\hat{\hat{\beta}} = \beta_o + \frac{1}{m}\,\Pi'U$$

where $\hat{\hat{\beta}} = \frac{1}{m}\Pi'\hat{\beta}$ with $\hat{\beta}' = \left[\hat{\beta}_1'\hat{\beta}_2'\ldots\hat{\beta}_m'\right]$, $\Pi' = [I \ I \ldots I]$, and $U' = [U_{1o}'U_{2o}'\ldots U_{mo}']$. Accordingly, vector $\hat{\hat{\beta}}$ contains the arithmetic average of the prior estimates $\hat{\beta}_i$.

We can now construct an expanded model (Theil 1971, p. 671)

$$\begin{bmatrix} y_o \\ \\ \hat{\hat{\beta}} \end{bmatrix} = \begin{bmatrix} X_o \\ \\ I \end{bmatrix} \beta_o + \begin{bmatrix} u_o \\ \\ \frac{1}{m}\Pi'U \end{bmatrix}$$

or, simply, $y_o^* = X_o^*\beta_o + u_o^*$ with the variance-covariance matrix of u_o^* equal to

$$E(u_o^* u_o^{*\prime}) = \Omega_o = \begin{bmatrix} \sigma_o^2 I & 0 \\ 0 & \frac{1}{m^2}\left(\Pi'\Omega\Pi\right) \end{bmatrix}.$$

Note that

$$\Pi'\Omega\Pi = \Pi' \begin{bmatrix} \Sigma_1^* & \Sigma & .. & .. & \Sigma \\ \Sigma & \Sigma_2^* & .. & .. & \Sigma \\ . & . & . & . & . \\ . & . & . & . & . \\ \Sigma & \Sigma & .. & .. & \Sigma_m^* \end{bmatrix} \Pi$$

with Σ_i^* for i = 1, 2, ..., m as defined in (5). Hence,

$$\Pi'\Omega\Pi = m(m-1)\Sigma + \sum_{i=1}^{m} \Sigma_i^*.$$

Assuming σ_o^2 and Σ are known and using the extended Gauss-Markov theorem, we obtain the estimator

$$\hat{\beta}_o = (X_o^{*\prime}\Omega_o^{-1}X_o^*)^{-1}X_o^{*\prime}\Omega_o^{-1}y_o^*. \qquad (7)$$

Following Leamer (1978, p.82), $\hat{\beta}_o$ can be expressed as a matrix weighted average of the OLS estimate of β_o (given X_o is full rank) and $\hat{\hat{\beta}}$ in (6). The empirical Bayes estimate assuming normality (i.e. G=N in (3)) which Aigner and Leamer (1984) derive is a matrix weighted average of the same OLS estimate and $\hat{\hat{\hat{\beta}}}$, where the latter is in itself a matrix weighted average of $\hat{\beta}_i$ for i = 1, 2, ..., m. As $\hat{\hat{\beta}}$ is the midpoint of the line segment joining the $\hat{\beta}_i$'s (Chamberlain and Leamer 1976), Vanhonacker and Price (1992) conclude that $\hat{\beta}$ must be at the centroid of all posterior means $\hat{\hat{\beta}}$.

3. <u>Efficiency and Precision of the Derived Estimates</u>

The variance-covariance matrix of $(\hat{\beta}_o - \beta_o)$ equals $Var(\hat{\beta}_o - \beta_o) = [X_0^{*}{}'\Omega_0^{-1}X_o^{*}]^{-1}$ which can be written alternatively as

$$Var(\hat{\beta}_o-\beta_o) = [\frac{1}{\sigma_0^2} X_o'X_o + m^2(m(m+1)\Sigma + \sum_{i=1}^{m} \sigma_i^2 (X_i'X_i)^{-1})^{-1}]^{-1} \qquad (8)$$

where the terms are defined as above. Since $\sigma_i^2 (X_i'X_i)^{-1}$ equals the variance-covariance matrix of the prior estimate $(\hat{\beta}_i-\beta_i)$ and this for i = 1, 2, ..., m, expression (8) shows that the efficiency of the recursive estimator in (7) is a function of (a) the <u>sample precision matrix</u> of the null situation (i.e., $X_O'X_O$), (b) the <u>potential for transfer</u> as captured in matrix Σ, and (c) the <u>efficiency of the prior estimates</u>. The latter two components contribute to potential gain in efficiency due to data transfer over the efficiency which could be achieved if β_O in model (1) were estimated independently. If the prior estimates are not very efficient, no matter how much transfer there is, nothing will be gained in the end. If there is no transfer, no matter how efficient the prior estimates are, the final estimates and, hence, their precision are independent from the priors. Algebraically, if either Σ or $(X_i'X_i)^{-1}$ for i = 1, 2, ..., m converges to an infinite matrix, the variance-covariance matrix in (8) reduces to

$$Var(\hat{\beta}_o-\beta_o) = \sigma_0^2(X_o'X_o)^{-1} \qquad (9)$$

which would be the variance-covariance matrix of the OLS estimate of β_O in model (1) using only the information contained in that model (and, of course, X_O being of full rank). This case is referred to as <u>complete non-transferability</u> and expression (9) forms a lower bound on the efficiency.

The upper bound on efficiency is reached in the case of complete transferability (i.e., when $\Sigma = 0$). In this instance, the variance-covariance matrix in (8) reduces to

$$\mathrm{Var}(\hat{\beta}_o-\beta_o) = \left[1/\sigma_o^2(X_o'X_o) + m^2\left(\sum_{i=1}^{m} \sigma_i^2\,(X_i'X_i)^{-1}\right)^{-1}\right]. \qquad (10)$$

For any matrix Σ between an infinite matrix and a zero matrix, the efficiency of the recursive $\hat{\beta}_o$ estimate will be between expression (9) and expression (10).

Two observations are in order. First, the variance-covariance matrix in (9) cannot be computed in our problem because X_o is not a full rank matrix. We can derive its inverse, however, which is the precision matrix which will form a lower bound on the precision with which vector β_o can be estimated. Second, in the case of complete data transferability (i.e., when $\Sigma = 0$), the density in (3) collapses to the mean $\bar{\beta}$ and all models can be pooled and estimated simultaneously as $\beta_i = \beta_o = \bar{\beta}$ (for i = 1, 2, ..., m). This pooled estimate of $\bar{\beta}$ (and, hence, β_o) will be the most efficient estimate with variance-covariance matrix

$$\mathrm{Var}(\hat{\beta}_o-\beta_o) = \left[(1/\sigma_o^2)(X_o'X_o) + \sum_{i=1}^{m} (1/\sigma_i^2)(X_i'X_i)\right]^{-1}. \qquad (11)$$

Comparing expressions (10) and (11), it is clear that the recursive estimator (7) is *not* the most efficient estimator in the event of complete data transferability. This is a direct result of the independent estimation of the β_i vectors (for i = 1, 2, ..., m) which form the prior information in the recursive estimator (Vanhonacker and Price 1992). In contrast, the pooled estimator whose variance-covariance matrix forms the upper bound on efficiency shown in (11) estimates all parameters simultaneously with the implicit constraint that the parameters are identical across situations. Accordingly, by estimating the m cases in (2) one by one, not incorporating any constraints on the parameters, there is a loss in efficiency when the recursive estimate $\hat{\beta}_o$ is derived, despite complete transferability.

4. Measuring the Amount of Transferability

The amount of realized data transfer, τ_o, can be measured as (Aigner and Leamer 1984):

$$\tau_o = |Var(\widehat{\beta}_o-\beta_o)^{-1}| / |(1/\sigma_o^2)(X_o'X_o) + \sum_{i=1}^{m}(1/\sigma_i^2)(X_i'X_i)|. \qquad (12)$$

In words, the amount of data transfer equals the determinant of the precision matrix of the recursive estimate $\widehat{\beta}_o$ over the determinant of the precision matrix of the pooled estimate. Note that τ_o can be computed at the individual-parameter level by taking the ratio of the corresponding diagonal elements in the matrices defined in the numerator and denominator of (12).

As the amount of data transfer increases, τ_o increases. The maximum value of τ_o is less than one, however, because the $\widehat{\beta}_i$'s for i = 1, 2, ..., m are estimated separately and there is a loss in efficiency relative to the pooled estimate as discussed above. From (10), the upper bound on τ_o equals

$$\max \tau_o = |[\frac{1}{\sigma_o^2}X_o'X_o + m^2[\sum_{i=1}^{m}\sigma_i^2(X_i'X_i)^{-1}]^{-1}]| / |[\frac{1}{\sigma_o^2}X_o'X_o + \sum_{i=1}^{m}\frac{1}{\sigma_i^2}(X_i'X_i)]|. \qquad (13)$$

Another observation about τ_o in (12) concerns its interpretation. Although Aigner and Leamer (1984) proposed the index as a measure of realized transfer, by definition the index measures the fraction of the pooled estimator's efficiency that is realized by the recursive estimator. Accordingly, any element that favorably affects the relative efficiency of the pooled estimate will drive the τ_o measure up irrespective of the amount of actual transfer. For example, an increase in residual variance of the prior studies pushes the τ_o measure up. This may seem counter-intuitive given expression (8), which shows that lower efficiency of the prior estimates implies lower efficiency of the recursive estimator and, hence, lower values of τ_o. An algebraic proof is straightforward, however, if one considers the case $\sigma^2 = \sigma_i^2$ for i = 0, 1, 2, ..., m. Increased residual variance does not mean more transfer but simply that a pooled estimate would become much more efficient than the recursive estimate. Hence, the Aigner and Leamer (1984) measure is useful as an indicator of relative gain in efficiency but should not be strictly considered a measure of the exact amount of transfer. In order to better understand the behavior of this measure, and the behavior of the recursive estimator itself, both were studied in a numerical simulation environment.

III - NUMERICAL SIMULATION RESULTS

1. Design

A simulation study was performed to assess the performance of the τ_o measure and the recursive estimator for data transferability. Parameter vectors for a model containing four parameters (zero intercept) were generated out of a multivariate normal distribution with mean zero and a variance-covariance matrix with diagonal elements equal to σ_β^2 and off-diagonal elements equal to $\rho\sigma_\beta^2$ (i.e., ρ measures the correlation between each pair of parameters). Eleven vectors were drawn randomly, ten for the prior studies and one for the new situation. According to Breusch (1980), the particular values of the parameters do not have an impact on the findings (see also Vanhonacker 1984).

The observations for the four predictor variables were generated out of a multivariate normal distribution with mean zero and a diagonal variance-covariance matrix with elements equal to σ_X^2. To each of the predictor values generated, a constant was added that equalled N.k where k denotes the column of the data matrix corresponding to that predictor. The disturbances were drawn randomly out of a multivariate normal distribution with mean zero and a diagonal variance-covariance matrix with elements equal to σ_u^2. Except for N and σ_X^2, all parameters of the design were varied as shown in Table 1.

Variations in σ_β^2 and ρ correspond to variations in Σ and, hence, pertain to the potential for information transfer as defined in (3). Ideal conditions for transferability are high parameter correlation, ρ, and relatively low parameter variance, σ_{β}^2. Variations in σ_u^2, on the other hand, affect the effiency of the prior estimates $\hat{\beta}_i$, which directly affects the efficiency of the recursive estimator in (7). As discussed above, however, σ_u^2 also affects the relative efficiency of the recursive estimator versus the pooled estimator. It thus is possible that an increase in σ_u^2 would lead to a decrease in the efficiency of the recursive estimate but an increase in τ_0.

Manipulation of these 3 parameters led to a 27 cell design. In each of the cells, 50 replications for the predictors and the disturbances were performed (i.e., the eleven parameter vectors were kept constant over replications in each cell). All values were obtained using the GGNSM routine in IMSL (1980).

Within this framework, two sets of transferability results were derived. In one instance, observations were assumed to be missing for the fourth predictor, whereas in the other instance, observations were assumed to be missing for the last two predictors. In other words, transfer into a single parameter as well as into a pair of parameters in a four parameter specification was assessed.

2. Simulation Results

The mean squared error (MSE) results are summarized in Tables 2 and 3, where Table 2 shows MSE for all four parameters and Table 3 shows MSE for only those parameters corresponding to missing data. ANOVA results on the MSE values are shown in Table 4.

Not surprisingly, the results in Table 2 indicate that MSE increased with increased residual variance (σ_u^2), increased parameter variance (σ_{β}^2), and decreased correlation between the parameters ($\rho \rightarrow 0$). These results are corroborated by the ANOVA results tabulated in columns 1 and 2 of Table 4. As shown there ρ, σ_{β}^2, and σ_u^2 have highly

Table 1

Parameter Values of Simulation Study

Parameter	Values
σ_β^2 (diagonal elements of Σ in (3))	0.01, 0.10, 1.0
ρ (pairwise correlation of parameters in β)[a]	0.1, 0.5, 0.9
σ_u^2 (residual variance)	10.0, 100.0, 1000.0
σ_X^2 (variance of predictors)	1000.0
N (multiplicative factor for mean of predictors)[b]	10.0
m (number of prior studies)	10

[a] Off-diagonal elements of Σ in (3) were all equal to $\rho\sigma_\beta^2$.

[b] The predictor observations were generated out of a multivariate normal distribution with mean equal to N.k where k denotes the column of the data matrix corresponding to that predictor.

significant main effects. The cell means in Table 2 indicate that the significant pairwise interaction between ρ and σ_{β}^{2} can be interpreted as follows: as σ_{β}^{2} increased, MSE increased but the magnitude of the increase was less as ρ approached one. High levels of parameter dependence thus compensated to some extent for increased levels of "noise" introduced by σ_{β}^{2}, and transferability estimates remained relatively accurate.

Table 2 also shows that MSE increased as the number of variables with missing data increased. High levels of parameter correlation again diminished the negative influence of this factor as it had done for σ_{β}^{2}. Indeed, when $\rho = 0.9$, MSE remained relatively low, even when two of four variables were without observations. This makes sense, given that data for any one of the model's predictors would contain a great deal of relevant information about the remaining predictors at such a high level of ρ.

For the instance where information about three parameters was transferred into a single fourth parameter (Column 2, Table 4), the ANOVA results also indicate significant pairwise interactions between residual variance σ_{u}^{2} and parameter variance σ_{β}^{2}, and between residual variance (σ_{u}^{2}) and correlation between parameters (ρ). These interactions enhance the effects in the direction of the corresponding main effects.

Table 3 summarizes the MSE results for the transferred parameters. In other words, the MSE for the parameters for which data were available in the new situation (i.e., parameters corresponding to variables with non-zero columns in X_O) are not incorporated. The tabulated MSE values are larger than the corresponding values in Table 2, but the pattern of results is similar. Except for an insignificant main effect of residual variance, these results are confirmed by the ANOVA results in columns 3 and 4 of Table 4. In sum, the findings of this simulation suggest that information transfer will be high, and parameter estimates will be accurate when the parameters are highly correlated but have small variances, and few predictors have non-zero observations (i.e., the number of parameters into which information is transferred is small relative to the total number of parameters). Moreover, when parameter correlations are very high, transferability results can be accurate, even when parameter variance is high and relatively few variables have data. When correlations are more moderate, however, these other two factors have a strong negative effect on the accurancy of the transferred parameter estimates. Interestingly enough, residual variance is less of an issue as exemplified by the ANOVA results on the transferred parameters.

Figure 1 illustrates the results obtained for the parameter-level τ_0 measure corresponding to (12). ANOVA results (not shown here) indicated that all main and interaction effects were significant at least at the 0.05 level. Consistent with the above discussion, τ_0 increased with residual variance (σ_u^2), particularly when the parameter variance (σ_β^2) was small. The measure is inversely related to parameter variance (σ_β^2). τ_0 also increased with increased parameter correlation (ρ). In sum, apart from the residual variance, the variables which drive τ_0 coincide with those which were identified above as driving the accuracy of the transferred parameter estimates. The impact of residual variance is a direct result of the definition of the τ_0 measure as discussed above. This result reinforces the conclusion that the measure is difficult to interpret in an absolute sense of being a strict indicator of the amount of transfer.

Table 5 summarizes the corresponding results for the max τ_0 as defined in (13). Overall, the values indicate that approximately 11 to 14% of the potential efficiency which could be obtained from pooling was lost by using the recursive estimator. Interestingly enough, none of the main effects nor interaction effects were found to be significant in an ANOVA on the max τ_0 values. Hence, the loss seems to reflect only the loss in efficiency from sequential estimation. The near constant value of this ratio measure suggests that σ_β^2, σ_u^2, and ρ had very similar effects on the efficiency of both the pooled and the recursive estimates.

Table 2

MSE Results for All Parameters [a]

		Number of Transferred Parameters [b]					
		Two			One		
		Residual Variance (σ_u^2)			Residual Variance (σ_u^2)		
σ_β^2 [c]	ρ [d]	10	100	1000	10	100	1000
	0.1	0.0071	0.0073	0.0126	0.0048	0.0048	0.0115
0.01	0.5	0.0040	0.0041	0.0088	0.0028	0.0030	0.0086
	0.9	0.0008	0.0010	0.0073	0.0006	0.0008	0.0072
	0.1	0.0710	0.0746	0.0827	0.0473	0.0456	0.0583
0.10	0.5	0.0387	0.0400	0.0498	0.0292	0.0288	0.0371
	0.9	0.0081	0.0083	0.0161	0.0062	0.0063	0.0134
	0.1	0.7326	0.7027	0.7250	0.4551	0.4610	0.4956
1.00	0.5	0.3979	0.4131	0.4229	0.2891	0.3041	0.3054
	0.9	0.0812	0.0811	0.0962	0.0612	0.0611	0.0817

[a] Results are based on a model with 4 parameters and the number of prior studies (m) equal to 10.

[b] Number of transferred parameters means number of predictors for which data were missing.

[c] Variance of parameters (i.e., diagonal elements of Σ in (3)).

[d] Correlation of parameters (off-diagonal elements of Σ in (3) equal $\rho\sigma_\beta^2$).

Table 3

MSE Results for Transferred Parameters [a]

		Number of Transferred Parameters [b]					
		Two			One		
		Residual Variance (σ_u^2)			Residual Variance (σ_u^2)		
σ_β^2 [c]	ρ [d]	10	100	1000	10	100	1000
	0.1	0.0133	0.0136	0.0170	0.0186	0.0181	0.0216
0.01	0.5	0.0074	0.0076	0.0109	0.0111	0.0012	0.0107
	0.9	0.0016	0.0017	0.0062	0.0024	0.0027	0.0059
	0.1	0.1348	0.1392	0.1456	0.1837	0.1767	0.1872
0.10	0.5	0.0745	0.0755	0.0811	0.1154	0.1136	0.0994
	0.9	0.0154	0.0154	0.0209	0.0245	0.0246	0.0272
	0.1	1.3785	1.3288	1.3570	1.7671	1.7953	1.8795
1.00	0.5	0.7454	0.7809	0.7806	1.1459	1.1997	1.1473
	0.9	0.1541	0.1534	0.1656	0.2435	0.2427	0.2691

[a] Results are based on a model with 4 parameters and the number of prior studies (m) equal to 10.

[b] Number of transferred parameters means number of predictors for which data were missing.

[c] Variance of parameters (i.e., diagonal elements of Σ in (3)).

[d] Correlation of parameters (off-diagonal elements of Σ in (3) equal $\rho\sigma_\beta^2$).

Table 4

ANOVA Results for Mean Squared Errors:
Significance of F-Values[a]

	All Parameters		Transferred Parameters[b]	
	Number of Transferred Parameters			
	Two	One	Two	One
Main Effects[c]				
ρ	0.0001	0.0001	0.0001	0.0001
σ_β^2	0.0001	0.0001	0.0001	0.0001
σ_u^2	0.0018	0.0001	0.2281	0.1898
Interaction Effects				
$\rho \times \sigma_u^2$	0.2820	0.0931	0.2479	0.0655
$\rho \times \sigma_\beta^2$	0.0001	0.0001	0.0001	0.0001
$\sigma_u^2 \times \sigma_\beta^2$	0.7843	0.0010	0.9344	0.1178
$\rho \times \sigma_u^2 \times \sigma_\beta^2$	0.1527	0.2271	0.1270	0.2031
Overall Statistics				
F-value	1462.00	1375.51	1922.54	1250.66
Significance	0.0001	0.0001	0.0001	0.0001
R^2	0.9664	0.9643	0.9742	0.9609

[a] Results are based on a model with 4 parameters and the number of prior studies (m) equal to 10.

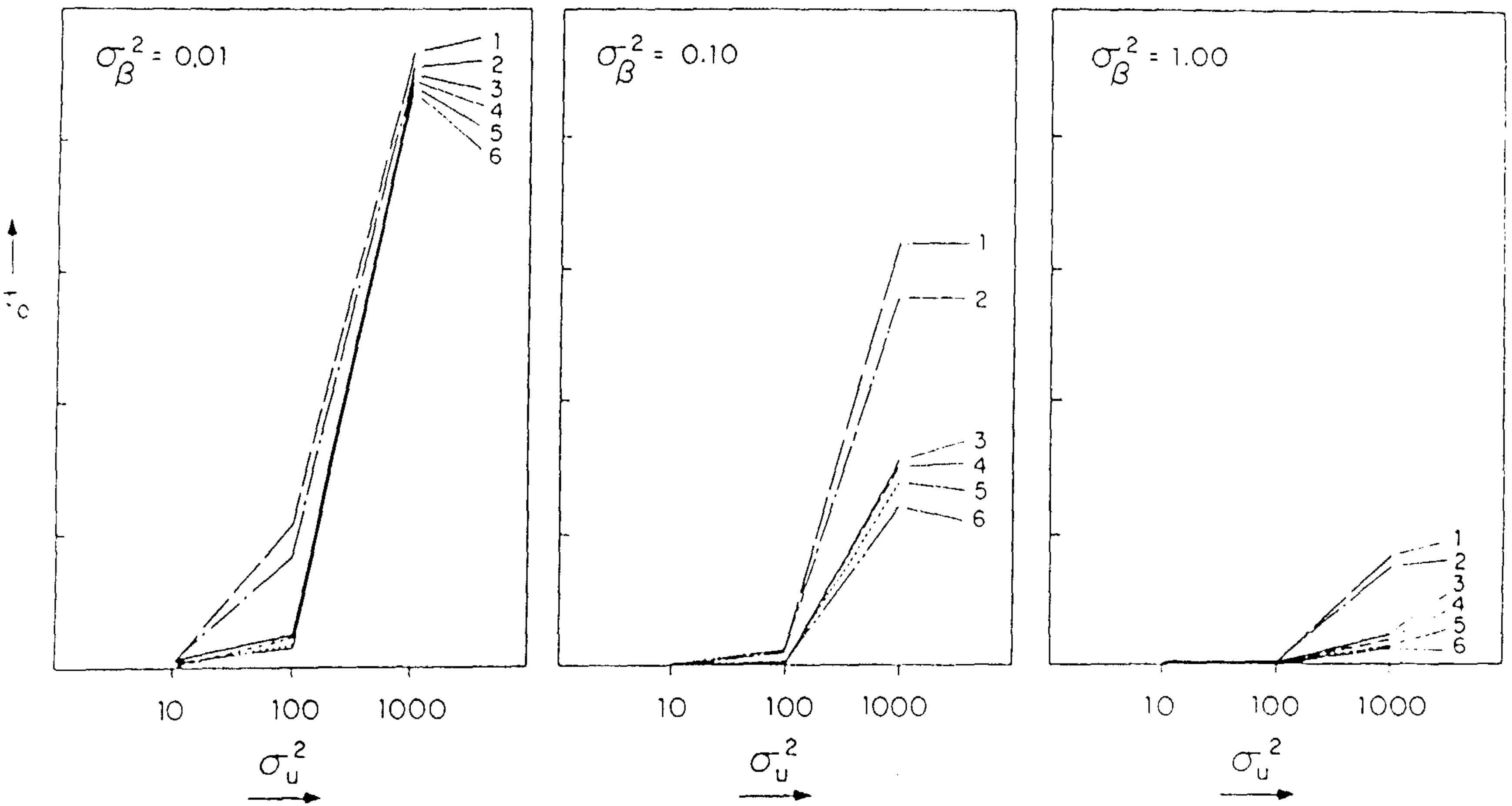
[b] Number of transferred parameters means number of predictors for which data were missing.

[c] ρ and σ^2 denote, respectively, the correlation and variance of the parameters (i.e., the diagonal elements of Σ in (3) equal σ_β^2 and the off-diagonal elements equal $\rho\sigma_\beta^2$; σ_u^2 denotes the residual variance.

IV - SUMMARY AND CONCLUSIONS

As practical instances do occur where the transfer problem arises, attention has been focused in the econometric literature on a formal methodology to derive estimates in the absence of data. Aigner and Leamer (1984) developed an empirical Bayes approach. Vanhonacker and Price (1992) developed a simpler recursive least-squares approach. The latter was discussed in some detail in this paper. The objective was to assess, within a numerical simulation environment, the efficiency of the recursive estimator and the amount of transfer which could occur.

The results indicate that efficient transfer estimates can be obtained when (a) relative to the total number of predictors, there are few with no observations, (b) efficient parameter estimates can be obtained for the historical cases, and (c) the parameters in the models are correlated. This latter factor is especially important. The residual variance seems to be less of an issue in deriving efficient estimates, but it does significantly affect the measure of transfer. Within the entire design, the loss in efficiency of the recursive estimate relative to the maximally efficient pooled estimate is less than 15%.



(1) Two-parameter transfer, $\rho = 0.9$
(2) One-parameter transfer, $\rho = 0.9$
(3) Two-parameter transfer, $\rho = 0.5$
(4) One-parameter transfer, $\rho = 0.5$
(5) Two-parameter transfer, $\rho = 0.1$
(6) One-parameter transfer, $\rho = 0.1$

Figure 1

Tau (τ_0) values

Table 5

MSE Results for Upper bound on τ_0 Measure (Max τ_0)[a]

		Two-Parameter Transfer[b] Residual Variance (σ_u^2)			Three-Parameter Transfer Residual Variance (σ_u^2)		
σ_β^2 [c]	ρ	10	100	1000	10	100	1000
	0.1	0.871491	0.870878	0.884772	0.871995	0.873372	0.884988
0.01	0.5	0.875183	0.878032	0.876932	0.875674	0.881563	0.881145
	0.9	0.875727	0.871949	0.880903	0.878709	0.868194	0.879811
	0.1	0.879287	0.870702	0.874137	0.876577	0.863595	0.874552
0.10	0.5	0.874245	0.878617	0.872821	0.878911	0.873244	0.872854
	0.9	0.873764	0.871545	0.871895	0.871190	0.874318	0.870292
	0.1	0.874496	0.87778	0.871349	0.878313	0.877614	0.873238
1.00	0.5	0.874298	0.867708	0.877274	0.873945	0.868556	0.875769
	0.9	0.878409	0.866475	0.879106	0.883301	0.869312	0.880528

[a] Results are based on a model with 4 parameters and the number of prior studies (m) equal to 10.

[b] Parameter-specific max τ_0 value computed as average of the two values corresponding to the two parameters.

[c] ρ and σ_β^2 denote, respectively, the correlation and variance of the parameters (i.e., the diagonal elements of Σ in (3) equal σ_β^2 and the off-diagonal elements equal $\rho\sigma_\beta^2$).

REFERENCES

Aigner, D. J. and E. E. Leamer (1984). Estimation of time-of-use pricing response in the absence of experimental data. *Journal of Econometrics* 26, 205-227.

Breusch, T.S. (1980). Useful invariance results for generalized regression models. *Journal of Econometrics* 13, 327-340.

Chamberlain, G. and E. E. Leamer (1976). Matrix weighted averages and posterior bounds. *Journal of the Royal Statistical Society* B38, 73-84.

Corstjens, M. (1991). Marketing Strategy in the Pharmaceutical Industry, Chapman and Hall, New York.

Duncan, D. B. and S. D. Horn (1972). Linear dynamic recursive estimation from the viewpoint of regression analysis," *Journal of the American Statistical Association* 67, 815-821.

IMSL Library, Edition 8 (1980). IMSL, Houston, Texas.

Leamer, E. E. (1978). Specification Searches. J. Wiley, New York.

Theil, H. (1971). Principles of Econometrics. J. Wiley, New York.

Vanhonacker, W.R. (1984). Finite sampling efficiency in a general linear model with serially correlated disturbances. *Journal of Statistical Computation and Simulation* 19, 185-203.

Vanhonacker, W. R. and L. J. Price (1992). Data transferability: estimating the response effect of future events based on historical analogy. Working Paper, INSEAD, Fontainebleau, France.