"DATA TRANSFERABILITY: ESTIMATING THE RESPONSE EFFECT OF FUTURE EVENTS BASED ON HISTORICAL ANALOGY"

by

Lydia PRICE*
and
Wilfried VANHONACKER**

N° 92/56/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
Data Transferability:
Estimating the Response Effect of Future Events
Based on Historical Analogy

Wilfried R. Vanhonacker
and
Lydia J. Price *

Revised September 1992

* Professor and Assistant Professor of Marketing, INSEAD, Fontainebleau, France. The authors would like to thank Mr Stephen C. Chappell of IMS and Dr J. Arversen of Pfizer for their help in providing the data for the empirical illustration. Funding was provided by the R&D Department of INSEAD. This research benefitted from critical comments by colleagues at Columbia University and INSEAD.
Data Transferability:
Estimating the Response Effect of Future Events
Based on Historical Analogy

ABSTRACT

Many managerial decisions are made with considerable uncertainty about their possible impacts. In assessing such problems, we often rely on analogous situations in the hope that they will shed light on the possible outcomes. This historical analogy assessment is often done in a rather ad hoc fashion. This paper introduces a formal methodology for transferring information or data from one situation to another. A simple recursive least squares approach to the Bayesian data transferability problem is discussed. Focus is on the method's applicability to commonly encountered marketing problems. Empirical testing and estimation issues are discussed and illustrated in the framework of generic substitution of brandname drug sales following patent expiration.

Marketing Science Index Key Words:
ECONOMETRIC MODELS; ESTIMATION AND OTHER STATISTICAL TECHNIQUES; FORECASTING; REGRESSION AND OTHER STATISTICAL TECHNIQUES.
1. Introduction

The effectiveness of managerial decisions and the impact of possible competitor actions often are assessed in the face of considerable uncertainty. Important data or information about the impact of the decision variable on some criterion either are missing, are not readily available, or are too expensive to obtain. Consider, for example, a local brand that has been available in the market for some time and has been able to secure a decent share position. Management now wants to enhance the brand’s appeal and is considering an advertising campaign to accomplish the task. Setting a campaign strategy is difficult, however, because the brand has never been supported by advertising. Can management, prior to its decision to advertise, evaluate the advertising responsiveness of the demand for its brand? Or consider a pharmaceutical product that has been able to build a strong brand image in the market and has been able to attain significant loyalty. Next year, its patent will expire and generic drugs are likely to enter the market with aggressive marketing programs. How will this change in competitive environment affect the product’s sales revenue? Finally, consider a company that has been experimenting with a new point-of-sale advertising program in a number of test markets. How effective will this program be when it is instituted in other crucial markets?

A reasonably accurate answer to each of these questions would be very valuable to management, who seeks to anticipate market responses and act (perhaps in a pre-emptive sense) to prevent any adverse effects on its market position. Obtaining such answers is not easy, however, as relevant data are missing. One obvious way to gain insights is to collect the relevant information in an experimental setting. But this approach could be expensive or its value could be questionable. Furthermore, time constraints might make it impossible to engage in such research.

Despite the diversity of the problems mentioned above, all can be characterized by one important feature: management is not totally ignorant about possible effects. Some information is available. For the local brand that is considering an advertising campaign, there is information on other brands (perhaps local ones) that have advertised in the past. For the drug that is coming off patent, management is familiar with other drugs that have come off patent faced with competitive action. The point-of-sale program has been tested in a number of areas, providing some data on its effectiveness. Can this information help management to assess the effectiveness of the pending decision or anticipated environmental change? Fortunately, the answer to this question is yes.

Historical analogies have long been used in qualitatively forecasting the outcomes of events for which no data are available (see, e.g., Armstrong 1978, Wheelwright and Makridakis 1977). Here, the new event is matched with similar historical events and,
depending on the degree of similarity, known outcomes of the historical events are used to develop a subjective forecast of outcomes for the new event.

A natural extension of this qualitative forecasting approach is to use quantitative forecasting models that have been validated in one context to predict the outcomes of events in another context. Researchers in the transportation field have investigated the extent to which empirically validated choice models can be used across contexts that vary in time frame or geographical location (see, e.g., Atherton and Ben-Akiva 1976, Koppelman and Rose 1983, Koppelman and Pas 1986, and Koppelman and Wilmot 1986). The assumption typically made in this research is that model parameters remain constant across contexts, with the exception of a situation-specific constant. Interesting empirical results have been established in this way, but no theory has been developed and the work falls short of providing a general but rigorous analytical framework to perform information transfer.

Recent research in econometrics provides a more formal framework and solution methodology for information transfer. Aigner and Leamer (1984) proposed a general analytic framework with a strict Bayesian solution methodology. The approach seeks to update the parameters of an empirically validated model from one context with limited information from a new context under the assumptions that all model parameters vary across contexts and are related within contexts. The procedure exploits knowledge about parameter dependency to transfer information from one context to another. Using this methodology, the answers to our questions can be inferred from transferred information. If the transfers are sufficiently successful, valuable insights can be obtained without any additional expense.

The solution methodology provided by Aigner and Leamer (1984) is based on a number of strict assumptions. Although theoretical arguments are provided, their work falls short of providing a full empirical account of the stringency of these assumptions and how they might be tested, and of how the methodology could be implemented given common constraints on available data. Furthermore, the empirical Bayes character of their solution methodology does not appeal to the novice user. Nevertheless, as the methodology does provide an innovative way to gain additional insight into difficult managerial issues, its potential applicability to marketing problems should be investigated.

The objective of this paper is to provide an overview of the transfer methodology so as to make it accessible to applied researchers in marketing. A common recursive least squares representation of the empirical Bayes approach is advocated in order to make it more comprehensible without sacrificing on statistical properties and rigour. With the focus on implementation, special attention is devoted to the various empirical questions arising in executing the methodology.
The paper is outlined as follows. First, the econometric interpretation of the problem of data transferability is briefly reviewed with special focus on the specific assumptions underlying the solution methodology. Second, the recursive least squares representation is described. Third, a discussion of the implementation task is provided with focus on various empirical questions that arise. Fourth, an empirical illustration is discussed in detail. Finally, conclusions are drawn on the general applicability of the transfer methodology.

2. The Problem of Data Transferability

Following Aigner and Leamer (1984), the problem of data transferability can be cast in an econometric framework. The procedure requires that a number of situations or cases are known for which historical observations are available. For the drug that goes off patent, for example, we have time-series observations on sales and a set of predictor variables collected for other drugs since they lost patent protection. In addition to uncontrollable environmental variables, those predictor variables might contain marketing mix variables of the company and its competitors (generic competitors as well as other brandname competitors whose drugs are promoted for the same indication in the common therapeutic class).

Accordingly, we can specify a set of sales response models, one for each of the drugs. Using matrix algebra notation, these models can be expressed as

\[ y_i = X_i \beta_i + u_i \quad \text{for } i = 1, 2, ..., m \]  

where \( y_i \) is a vector of dimension \((n \times 1)\) that contains the sales observations for drug \( i \), \( X_i \) is a matrix of dimension \((n \times k)\) that contains observations on \( k \) predictor variables, \( \beta_i \) is a vector of dimension \((k \times 1)\) that contains the response parameters, and \( u_i \) is a vector of dimension \((n \times 1)\) that contains random disturbances. The \( n \) observations on each variable are time series starting at the patent expiration date. As the response environment during that time period is different from the environment prior to patent expiration, some valid predictors will be contained in matrix \( X_i \) which did not play any role in sales response when the drug was still under patent protection (e.g., price of generic competitors). Other predictors are reasonable both before and after patent expiration (e.g., price of drug itself). The first \( l \) predictors in matrix \( X_i \) \((l < k)\) are common predictors (i.e., common to both time periods); the remaining \((k - l)\) predictors are only appropriate and valid for post-patent-protection sales response. Partitioning the parameter vector \( \beta_i \) accordingly, (1) can be expressed as

\[ y_i = X_{i1} \beta_{i1} + X_{i2} \beta_{i2} + u_i \quad \text{for } i = 1, 2, ..., m \]
where matrix $X_{i1}$ is $(n \times l)$ and contains observations on the common predictors, and matrix $X_{i2}$ is $[n \times (k - l)]$ and contains observations on the post-patent-protection predictors.

With respect to the historical response models in (1) or (2), a number of basic technical assumptions are made. All predictor variables are non-stochastic. For notational simplicity, we will assume that an equal number of time-series observations are available for the variables in each model in (1). The methodology is not constrained to such a case, as will become evident. The approach will require, however, that there are enough observations for each of the m historical cases so that independent estimates of $\beta_i$ for $i = 1, 2, ..., m$ can be obtained. Those independent estimates constitute the information that will be transferred into the new situation.

The parameter vectors $\beta_i$ in (1) will be different for the individual cases. The response model specifications are assumed to be structurally identical, however. Specifically, they are assumed to have identical functional forms (e.g., linear, multiplicative, etc.) and to contain an identical set of predictor variables. Continuing our example, the assumption stipulates that each drug has an identical model describing its sales response since it lost patent protection. If, for example, the functional form of the response model is multiplicative for some drugs and linear for others, they do not represent analogous events and nothing can be learned about one drug by observing the other drugs within the framework adopted here. This basic assumption of structural equality is one aspect of historical "analogy".

For our drug that will come off patent next year, we have a similar response model describing anticipated post-patent-protection sales response. Specifically,

$$y_o = X_o \beta_o + u_o$$

(3)

with terms having definitions identical to those in (1). Accordingly, the structural equality goes beyond the historical cases to include the new situation. This assumption is another aspect of historical "analogy".

As we have been selling the drug for a number of years, we have time-series observations on the first $l$ predictors specified in model (3). Because of patent protection, we have not as yet faced any direct generic competitors and, hence, observations on generic drug prices or other competitive variables which are likely to impact sales after the expiration date are lacking. Consistent with notation adopted above, we as yet have no observations on the last $(k - l)$ predictors in matrix $X_o$. To apply the transferability methodology, the columns of $X_o$ corresponding to those predictors will contain zero entries. Because of these zero entries, the matrix is not full rank and no independent least squares estimate can be directly obtained for $\beta_o$.
in (3). The data transferability methodology will attempt to estimate $\beta_0$ from estimates for $\beta_i$ ($i=1,2,...,m$). These latter estimates include values for the last $(k - 1)$ parameters corresponding to predictors that will play a role in future sales response. An assessment of the likely impact of these predictors informs us on the response environment we are likely to face and enables us to adapt predictors presently under our control in order to improve our competitive position in the future. In our pharmaceutical illustration, for example, management might use information about sensitivity to generic price competition in order to adjust their own pricing strategies before the generics enter the market. Prior to further discussion of the estimation and forecasting procedure, however, an additional aspect of historical analogy has to be specified.

Structural equivalence in response models does not by itself justify drawing inferences about response parameters across situations. The parameters have to be linked in some manner which enables transfer of information or knowledge. As another aspect of historical analogy, therefore, it is assumed that there is a link between the parameters $\beta_i$ for $i = 1,2,...,m$ and $\beta_0$. These links are incorporated in an econometric model using a random coefficients framework. Stochastic parameter variation has some tradition in sales response modeling (for an overview see Hanssens, Parsons, and Schultz 1990, pp. 59-61), and the transferability approach discussed here extends this work.

In data transferability, it is assumed that the vectors of response coefficients are distributed around some mean vector $\bar{\beta}$ with a variance-covariance matrix $\Sigma$, or

$$\beta_i \sim G(\bar{\beta}, \Sigma) \quad \text{for } i = 0,1,2,...,m \quad (4)$$

where $G$ denotes a multivariate density. Aigner and Learner (1984) assume $G$ to be the multivariate normal density so they can rely on the classic Lindley and Smith (1972) results to derive their empirical Bayes estimator. As we will rely on the common recursive least squares representation of Bayesian regression, no particular distributional assumptions have to be made with respect to $G$. Nevertheless, statistical inferencing will require a specific multivariate density and, as most inference tests require normality, we will fall back on this assumption for statistical post-estimation evaluation.

Note that consistent with (2), we can partition the variance-covariance matrix $\Sigma$ as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix} \quad (5)$$
where the off-diagonal matrix \( \Sigma_{12} \) contains the covariances between \( \beta_{i1} \) and \( \beta_{i2} \) (for \( i = 1,2,...,m \)). This matrix plays a crucial role in the transferability methodology as it provides the link between the two sets of parameters. Matrix \( \Sigma \) also provides the link between the response coefficients of various instances. If matrix \( \Sigma \) is a null matrix, then the multivariate density in (4) collapses to the mean value \( \bar{\beta} \) and all regression models in (1) and (3) have identical parameters. Accordingly, the most efficient estimate of \( \beta_0 \) in (3) can be obtained from pooling the various time series (Bass and Wittink 1975). If by contrast \( \Sigma \) converges to an infinite matrix, then efficient estimates of the response coefficients can only be obtained from estimating each model in (1) and (3) separately. Note that the latter is true irrespective of the size of the covariances (i.e., the off-diagonal elements of \( \Sigma \) ). Non-zero covariances indicate a potential for transfer, but infinitely large variances give rise to highly inefficient estimates. In other words, a great deal of noise would be transferred, which would not enhance the efficiency of the final estimates.

Structural equivalence of the response models in (1) and (3) together with parameter dependence in a random coefficients framework constitute the operational definition of historical analogy. As the transferability methodology is only applicable to instances where historical analogy is present, these two conditions will have to be established before the transfer can be attempted. How this is done is discussed extensively later. We first proceed with the theoretical development of the transferability methodology itself.

### 3. Data Transferability: Recursive Least Squares Approach

Since historical data are available for all variables specified in (1), estimates can be obtained for the parameter vectors \( \beta_i \) for \( i = 1,2,...,m \). Assuming that \( \beta_i \) is independent of \( u_i \), we can rely on the extended Gauss-Markov theorem to obtain the identical result of the fixed nonrandom regression problem (see, e.g., Duncan and Horn 1972). In other words, under the normal assumptions that the disturbances in \( u_i \) are independently and identically distributed with a mean zero and a variance-covariance matrix \( \sigma^2 I \), the OLS estimator of \( \beta_i \), \( \hat{\beta}_i = (X_i'X_i)^{-1} X_i'y_i \), is the "best linear unbiased estimator" (BLUE). \(^1\)

Consistent with (4), we can write that

\[
\beta_i = \bar{\beta} + e_i \quad \text{for } i = 0,1,2,...,m
\]

with \( E(e_i) = 0 \) and \( E(e_i'e_i) = \Sigma \) where \( E \) denotes the expectations operator. This expression will be useful in subsequent derivations.
Furthermore,

$$\beta_i - \beta_0 = e_i - e_0 \quad \text{for } i = 1, 2, ..., m$$

or

$$\beta_i = \beta_0 + V_{io} \quad \text{for } i = 1, 2, ..., m$$

where $V_{io} = e_i - e_0$. Incorporating the $\hat{\beta}_i$ estimates discussed above, it can be shown that (see Appendix 1)

$$\hat{\beta}_i = \beta_0 + U_{io} \quad \text{for } i = 1, 2, ..., m \quad (7)$$

where $U_{io} = (X_i' X_i)^{-1} X_i' u_i + V_{io}$. Accordingly, expression (7) states that the least squares estimates $\hat{\beta}_i$ for $i = 1, 2, ..., m$ can be interpreted as prior estimates of $\beta_0$, the parameter vector of model (3).

The prior information on $\beta_0$ can now be updated with the sample information available for regression model (3) using a Goldberger-Theil approach (Theil 1971, p. 671). We discuss this approach assuming that all random elements introduced above are independent from one another. This assumption is quite restrictive but inconsequential to the basic methodology. Under dependence, variance-covariance matrices of disturbances would be complex, full variance-covariance matrices. We make the assumption of independence here to simplify the notation and as such give a more insightful and intuitive representation of the approach.

The recursive least squares representation (or Goldberger-Theil approach) is based on expanding model (3) with the prior information. That information can be integrated in two ways: either the parameter vector for each historical case is considered as an individual prior using (7), or these parameter vectors are combined (e.g., arithmetic mean) to form a single prior. In the first instance, the expanded model becomes

$$\begin{bmatrix}
y_0 \\
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\vdots \\
\hat{\beta}_m
\end{bmatrix} = \begin{bmatrix}X_0 & I & U_0 \\
I & I & U_{10} \\
I & I & U_{20} \\
\vdots & & \vdots \\
I & I & U_{mo}
\end{bmatrix}$$

(8)
with $U_{io}$ for $i = 1, 2, \ldots, m$ as defined in (7).

In the second instance, the expanded model becomes

$$
\begin{bmatrix}
y_o \\
\beta
\end{bmatrix} =
\begin{bmatrix}
X_o \\
I
\end{bmatrix} \beta_0 +
\begin{bmatrix}
u_o \\
\frac{1}{m} \Pi' U
\end{bmatrix}
$$

(9)

where $\hat{\beta} = (1/m) \Pi' \hat{\beta}$ with $\hat{\beta} = [\hat{\beta}_1 \hat{\beta}_2 \ldots \hat{\beta}_m]$, $\Pi' = [I I \ldots I]$, and $U' = [U'_{1o} U'_{2o} \ldots U'_{mo}]$.

The recursive least squares approach to data transferability amounts to applying the extended Gauss-Markov theorem to either (8) or (9) to derive a Generalized Least Squares (GLS) estimate for parameter vector $\beta_0$. For the individual priors (expression (8)), it can be shown that (Learner 1978, p. 82)

$$
\hat{\beta}_0 = \left[\frac{1}{\sigma_0^2} X_o' X_o + \Pi' \Omega^{-1} \Pi \right]^{-1} \left[\frac{1}{\sigma_0^2} X_o' y_o + \Pi' \Omega^{-1} \hat{\beta} \right]
$$

(10)

with

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

where $\Sigma_i^* = E(U_{i0}U_{i0}') = 2\Sigma + \sigma_i^2 (X_i'X_i)^{-1}$.

For the combined priors (expression (9)), it can be shown that

$$
\hat{\beta}_0 = \left[\frac{1}{\sigma_0^2} X_o' X_o + m^2 (\Pi' \Omega \Pi)^{-1} \right]^{-1} \left[\frac{1}{\sigma_0^2} X_o' y_o + m^2 (\Pi' \Omega \Pi)^{-1} \hat{\beta} \right].
$$

(11)

Except in some special circumstances (e.g., when $m=1$), the recursive estimates in (10) and (11) are not identical. In the case of the individual priors, the weight of each prior parameter vector $\hat{\beta}_i$ ($i=1, 2, \ldots, m$) is proportional to the efficiency with which it was estimated; in the case of the combined - arithmetic mean - prior, the weight of the prior (to which each prior parameter vector contributes in an identical fashion) is proportional to the average efficiency of the individual priors. Hence, if a single prior is relatively inefficient, only that particular prior
will be discounted in \( \hat{\beta}_0 \) in (10). However, in the combined prior case, all priors will be somewhat discounted. Given that the transfer is sensitive to prior efficiency (Price and Vanhonacker 1992), and for obvious degree-of-freedom reasons, it seems preferable to work with the recursive estimator incorporating the individual priors (expression 10). In a different but related approach, Yamada and Leung (1992) provide empirical support for individual priors. It is interesting to note, however, that both recursive estimates are - under normality and quadratic loss - special cases of the empirical Bayes estimator derived by Aigner and Learner (1984). For a discussion, see Appendix 2.

The efficiency of the recursive estimator is captured by its variance-covariance matrix which equals the inverted matrix in brackets in expression (10) (and, hence, expression (11)). That expression shows that the efficiency of the \( \hat{\beta}_0 \) estimate is a function of (a) the sample precision matrix of the new situation (i.e., \( X_0'X_0 \)), (b) the potential for transfer as captured in matrix \( \Sigma \), and (c) the efficiency of the prior estimate. The latter two components contribute to potential gain in efficiency due to data transfer over the efficiency which could be achieved if \( \beta_0 \) in model (3) were, and could be, 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 \( \Sigma \) or \( (X_i'X_i)^{-1} \) for \( i = 1, 2, \ldots, m \) converges to an infinite matrix, the variance-covariance matrix of the recursive estimator reduces to

\[
\text{Var}(\hat{\beta}_0 - \beta_0) = \sigma_0^2(X_0'X_0)^{-1}
\]

which would be the variance-covariance matrix of the OLS estimate of \( \beta_0 \) in model (3) using only the information contained in that model (and, of course, \( X_0 \) being of full rank). Aigner and Learner (1984) refer to this case as complete non-transferability and expression (12) 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, all models can be pooled and estimated simultaneously since \( \hat{\beta}_i = \beta_0 = \bar{\beta} \) (for \( i = 1, 2, \ldots, m \)) and the density in (4) collapses to the mean \( \bar{\beta} \). The pooled estimate of \( \beta \) (and, hence, \( \hat{\beta}_0 \)) will be the most efficient estimate with variance-covariance matrix

\[
\text{Var}(\hat{\beta}_0 - \beta_0) = \left[ \frac{1}{\sigma_0^2} (X_0'X_0) + \sum_{i=1}^{m} \frac{1}{\sigma_i^2} (X_i'X_i) \right]^{-1}
\]
However, this upperbound on efficiency cannot be reached. Replacing $\Sigma$ with a zero matrix in the variance-covariance matrix of $\hat{\beta}_o$ does not provide the expression in (13). This is a direct result of the independent estimation of the $\beta_i$ vectors (for $i = 1, 2, ..., m$) which form the prior information in the recursive estimator. In contrast, the pooled estimator whose variance-covariance matrix forms the upper bound on efficiency estimates all parameters simultaneously with the implicit constraint that the parameters are identical across situations. Accordingly, by estimating the historical situations 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. Implementation: Some Empirical Issues

1. Some Estimation Issues

In deriving the estimator for $\hat{\beta}_o$ discussed above, we have assumed that $\sigma_o^2$ and $\Sigma$ are known. In practice, however, these will have to be estimated. A direct estimate of $\sigma_o^2$ using residuals of model (3) is impossible to obtain since $X_o$ is not of full rank in the problem investigated here. The estimate suggested here equals the sample average of the least squares estimates of $\sigma_i^2$ for $i = 1, 2, ..., m$. That is

$$\hat{\sigma}_o^2 = \frac{1}{m} \sum_{i=1}^{m} \hat{\sigma}_i^2$$  \hspace{1cm} (14)

where $\hat{\sigma}_i^2 = \left[ (\hat{y}_i - X_i \hat{\beta}_i) \right] \left( \hat{y}_i - X_i \hat{\beta}_i \right)/(n - k)$ for $i = 1, 2, ..., m$ and $\hat{\beta}_i$ are the OLS estimates of $\beta_i$ in (1).

The variance-covariance matrix of the random coefficients, $\Sigma$, can be estimated as follows

$$\hat{\Sigma} = \frac{1}{m} \sum_{i=1}^{m} \left( \hat{\beta}_i - \hat{\beta} \right) \left( \hat{\beta}_i - \hat{\beta} \right)'$$  \hspace{1cm} (15)

with $\hat{\beta}$ defined as above. Assuming normality, it can be shown that $\hat{\sigma}_o^2$ in (14) and $\hat{\Sigma}$ in (15) are related to the empirical Bayes estimators using non-informative priors on $\sigma_o^2$ and $\Sigma$. Using the classic Lindley and Smith (1972) results, it can be shown that $\hat{\sigma}_o^2$ is proportional to the empirical Bayes estimator and $\hat{\Sigma}$ is exactly identical to the empirical Bayes estimator. Note,
however, that when \( m \) is very small, the \( \hat{\Sigma} \) estimate shown in (15) might not be invertible (i.e., is numerically singular). In such a case, an empirical Bayes approach with an informative prior as suggested and illustrated in Aigner and Leamer (1984) might provide a good estimator for \( \Sigma \).

2. Testing Structural Equivalence and Parameter Dependence

The transferability methodology described above is grounded in two fundamental assumptions: structural equivalence and parameter dependence. Together, they constitute the operational definition of historical analogy. As the effectiveness of the methodology is entirely dependent on this definition, its appropriateness has to be assessed empirically before implementation. We will discuss here the various tests which have to be performed and their appropriate sequence.

Structural equivalence requires that each model have an identical set of predictor variables and an identical functional form. There are two ways to assess the validity of this assumption. The first is to estimate the best model for each historical case (where best is defined according to some apriori specified criterion or set of criteria) and then check to see that the predictor variables and functional form are the same across models. When only the model estimation results of the historical cases are available, this is the only procedure that can be followed. Although straightforward, this procedure is restrictive and it may lead to frequent rejection of the structural equivalence assumption. An alternative approach is possible when the original data are available. In this case, one specifies apriori structurally equivalent models. If these models are statistically significant, one can conclude that the equivalence assumption has been satisfied.

Comparing these two approaches, it is clear that each requires a tradeoff. The first approach is highly restrictive, but in the event of a positive outcome, the parameter estimates will be efficient. The second approach is less restrictive, and may yield a usable set of models, but at the risk of reduced parameter efficiency relative to the “best” model. This loss of efficiency is not without consequence, as it will influence the effectiveness of the subsequent transfer. Some loss of efficiency may be necessary, however, to assure that the structural equivalence assumption is met.

The second fundamental assumption that needs to be assessed is parameter dependence. The transferability methodology assumes that the true parameter vectors are drawn from a multivariate distribution (see expression (4)). Furthermore, the variance-covariance matrix of that distribution, \( \Sigma \), has to have a particular structure: \( \Sigma_{12} \) in (5) cannot be a zero matrix for if it were, the link for the transfer would be missing. Thus, to assure parameter dependence, it is
necessary to test that the parameter vectors come from the same multivariate density and that \( \Sigma_{12} \neq 0 \). However, testing these conditions is not simple. A problem arises from the fact that we do not directly observe sample points out of the multivariate density \( G \) (as the true parameter vectors are unknown). Instead, we have estimates of these sample points (the \( \hat{\beta}_i \)'s for \( i = 1, 2, \ldots, m \)) and some information on how well the estimates capture the sample points. Furthermore, the number of estimates is likely to be limited because of limited availability of structurally equivalent historical cases. Recognizing these caveats, we might cautiously rely on some known tests, assuming that \( G \) is a multivariate normal density and that the historical estimates are sample points out of that density.

Given a set of sample points, we can derive sample estimates for \( \hat{\beta} \) and \( \hat{\Sigma} \) as shown in (15). To determine whether the sample points are drawn from the same distribution, we can derive estimates from various subsets of the observations, and then test for parameter equivalence on a pairwise basis. Subsets of the sample observations can be chosen by a split-half approach (as in cross-validation), a subsampling process with replacement (as in bootstrapping), or a process by which sample points are sequentially dropped (as in jackknifing).

To test for the equality of subsample means (first moment), we can use the two-sample Hotelling \( T^2 \) statistic (F-test, Morrison 1976, p. 131). To test for the equality of the subsample variance-covariance matrices (second moment), we can use the chi-square test (Morrison 1976, p. 252). The power of these tests depends on the sample sizes and the appropriateness of the normality assumption. The latter is of particular concern for the chi-square test of covariance matrix homogeneity, because of the test's known sensitivity (Mardia 1971).

Finally, the test of parameter independence can be done directly on the sample estimate \( \hat{\Sigma} \) in (15). The null hypothesis of \( \Sigma_{12} = 0 \) can be assessed using either an F-test (when \( \Sigma_{12} \) is a vector of dimension \( (l \times 1) \); Morrison 1976, p. 108) or the largest-characteristic-root test (when \( \Sigma_{12} \) is a matrix; i.e., \( (k - l) > 1 \), Morrison 1976, p. 256). Again the power of these tests will depend on the sample sizes (i.e., number of historical cases) and the normality assumption.

3. Subset Transfer

The recursive least squares representation in (8) suggests the possibility of confining the transfer (and, hence, incorporation of prior information) to a subset of parameters. That is, one can exclude a subset of the common predictors in \( X_{i1} \) in (2) from the transfer process. Their corresponding parameters would not be included in \( \hat{\beta}_0 \) in (8), but would be estimated
solely on the basis of sample information for the new situation. For example, if the intercept is estimated freely and independently from the prior intercepts, the expanded model (analogous to (8)) would be

\[
\begin{bmatrix}
\gamma_o \\
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_m
\end{bmatrix} =
\begin{bmatrix}
e & X_o \\
o & I \\
o & I \\
\vdots & \vdots \\
o & I
\end{bmatrix}
\begin{bmatrix}
\beta_{oo} \\
\beta_o \\
\beta_o \\
\vdots \\
\beta_0
\end{bmatrix} +
\begin{bmatrix}
u_o \\
U_{10} \\
U_{20} \\
\vdots \\
U_{m0}
\end{bmatrix}
\]

where \( e \) denotes a column vector containing all ones. Scalar parameter \( \beta_{oo} \) denotes the intercept in the new case, where vector \( \beta_o \) (and all prior \( \beta_i \) for \( i = 1,2,...,m \)) contain only slope parameters.

The possibility of confining information transfer to a subset of parameters has a number of interesting practical implications. First, as the efficiency of the recursive estimator is influenced by the efficiency of the independently estimated priors (Price and Vanhonacker 1992), we could confine our analysis to the efficient priors in order to achieve the maximum possible gain from the transfer methodology. Second, subset transfer implies a possible relaxation of the structural equivalence assumption underlying the methodology. Essentially this implies that the equivalence restriction is confined to that part of the model incorporated in the transfer methodology. In other words, structural equivalence is required for all predictors in \( X_{2i} \) in (2) and for the subset of common predictors in \( X_{11} \) in (2) whose parameters are related to the former in a multivariate density sense as captured in (4).

4. **Analysis Steps**

Taking into account the empirical issues discussed above, the recursive least-squares approach to data transferability consists of the following steps:

(i) Estimate structurally equivalent models for the historical situations in (1) and obtain the parameter estimates (i.e., for \( i = 1,2,...,m \)), their variance-covariance matrices (i.e., \( \sigma^2_i(X_i'X_i)^{-1} \) for \( i = 1,2,...,m \)), and the variance of the residual errors (i.e., \( \sigma^2 \) for \( i = 1,2,...,m \));

(ii) Compute \( \sigma^2 \) as shown in (14) using the \( \sigma^2_i \)'s for \( i = 1,2,...,m \);
(iii) Compute the mean parameter vector, $\hat{\beta}$, across the historical situations and derive matrix $\hat{\Sigma}$ as shown in (15);

(iv) Test whether or not the $\hat{\beta}_i$ for $i = 1, 2, \ldots, m$ come from the same multivariate normal density using Hotelling's $T^2$ statistic (F-test) for the mean vector and the chi-square test for the variance-covariance matrix as discussed above;

(v) Test for parameter independence (i.e., $\Sigma_{12} = 0$ in expression (5)) using either the F-test (if $(k - l) = 1$) or the largest-characteristic-root test (if $(k - l) > 1$);

(vi) Construct the variance-covariance matrix $\Omega$ shown above using the results of steps 1, 2, and 3;

(vii) Derive the recursive estimator $\hat{\beta}_0$ as expressed in (10) or (11).

4. Empirical Illustration

The transferability methodology discussed above was applied to the antidepressant drug category. An overview of the various drugs in this therapeutic class along with their generic equivalents is shown in Table 1. The objective of the illustration was to assess sales response of a branded drug, Asendin, to price competition from generics entering the category upon patent expiration. The illustration demonstrates the transferability methodology, and also points out some of the issues that arise in its application. Asendin, manufactured by Lederle, went off patent in August 1989. Generic equivalents of Amoxapine, Asendin's generic substance, almost immediately came on the market. Four drugs in the antidepressant category have come off patent in recent years: Norpramin (July 1986), Sinequan (January 1986), Ludiomil (August 1985), and Adapin (January 1986). All four branded drugs had faced generic competition soon after their patents expired (on average, 12 months from the month in which the patent expired). The post-patent response environment for these drugs was assumed to provide information on Asendin's anticipated exposure to generic competition.

IMS Drugstore Audit data were available for the period from July 1984 until January 1990. The monthly time series contain for each drug "Extended Unit" sales and the average price paid by pharmacies. IMS provided data on retail markups for the time period to enable the derivation of average retail prices. The analyses focus on individual strength sizes of 10, 25, and 50 milligrams sold in packages of 100 units. These constitute the main revenue providers for antidepressant drugs. Note that Ludiomil and Asendin are only available in 25
Table 1

Anti-Depressant Drug Category a

<table>
<thead>
<tr>
<th>Brandname</th>
<th>Manufacturer</th>
<th>Date Introduced b</th>
<th>Expiration Date c</th>
<th>Generic Equivalent</th>
<th>Date Generic Appeared on Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elavil</td>
<td>Merck Sharp &amp; Dohme</td>
<td>4/61</td>
<td>5/76</td>
<td>Amitriptyline</td>
<td>10/77</td>
</tr>
<tr>
<td>Vivactil</td>
<td>Merck Sharp &amp; Dohme</td>
<td>11/67</td>
<td>3/85</td>
<td>Protriptyline</td>
<td>--</td>
</tr>
<tr>
<td>Norpramin</td>
<td>Merrell Dow</td>
<td>7/75</td>
<td>7/86</td>
<td>Desipramine</td>
<td>7/87</td>
</tr>
<tr>
<td>Pamelor</td>
<td>Sandoz</td>
<td>9/77</td>
<td>11/92</td>
<td>Nortriptyline</td>
<td>--</td>
</tr>
<tr>
<td>Sinequan</td>
<td>Roerig (Pfizer)</td>
<td>10/69</td>
<td>1/86</td>
<td>Doxepin</td>
<td>4/86</td>
</tr>
<tr>
<td>Tofranil</td>
<td>Geigy</td>
<td>1/59</td>
<td>5/68</td>
<td>Imipramine</td>
<td>5/76</td>
</tr>
<tr>
<td>Asendin</td>
<td>Lederle</td>
<td>10/80</td>
<td>8/89</td>
<td>Amoxapine</td>
<td>--</td>
</tr>
<tr>
<td>Ludiomil</td>
<td>Ciba</td>
<td>1/81</td>
<td>8/85</td>
<td>Maprotiline</td>
<td>2/88</td>
</tr>
<tr>
<td>Adapin</td>
<td>Pennwalt</td>
<td>2/73</td>
<td>1/86</td>
<td>Doxepin</td>
<td>4/86</td>
</tr>
<tr>
<td>Surmontil</td>
<td>Wyeth</td>
<td>10/79</td>
<td>5/68</td>
<td>Trimipramine</td>
<td>4/88</td>
</tr>
<tr>
<td>Aventyl</td>
<td>Lilly</td>
<td>12/64</td>
<td>11/92</td>
<td>Nortriptyline</td>
<td>--</td>
</tr>
</tbody>
</table>

a Category 64311 in the IMS directory, for the period July 1984 to January 1990.
b IMS Physicians Reference Desk Manual
c Obtained from Marketing Research Department at Pfizer
and 50 mg strengths. In total, 11 historical sales series were available representing post-patent-expiration market response for branded drugs of different strength sizes. In addition, 2 historical sales series were available for Asendin, representing pre-patent expiration market response.

An illustration of generic substitution in the antidepressant category is shown in Figure 1. The figure shows the extended unit sales for Sinequan and Adapin over time (i.e., the time leading up to patent expiration, subsequent generic entry, and the period of generic competition). Across all strength sizes, the sales dropped progressively following generic (Doxepin) entry. Because of the sizeable sales volume of these drugs, generic entry occurred within four months of the patent expiration date. Generic substitution was particularly strong in the higher strength categories and for Adapin whose 25 and 50 mg extended unit sales dropped 50% over a period of less than 2 years following generic introduction.

Figure 2 illustrates the price differential between Adapin, Sinequan, and their generic equivalent for the 25 mg strength size. The generic price is a weighted average across generic (Doxepin) manufacturers with extended unit sales being the weights. The large price differential is not atypical and explains the extensive generic substitution shown in Figure 1. For example, it appears that Adapin suffered more than Sinequan because of its premium pricing. Additionally, higher strength drugs suffered more, as they tend to be used in prolonged treatments where generic substitution would result in substantially lower treatment costs to the patient. This extensive generic substitution which apparently is driven by enormous price differentials, is the focus of the response modeling used to illustrate the transferability approach.

All extended unit sales series were adjusted for seasonality (monthly dummy variables) and cyclical trends (second degree polynomial) using the best linear unbiased (BLUE) seasonal adjustment suggested by Jorgensen (1964). These estimation results are not reported here but are available from the authors. The residual series were used to estimate apriori-specified, structurally-equivalent models. The models were specified taking into account the limited available data, the nature of the residual series after seasonal adjustment, and theoretical and empirical priors on sales response following patent expiration. As sales predictors, only prices were available. This is not as problematic as it may first appear, however, since prices are generally becoming more important in pharmaceutical marketing due to reimbursement policies and generic substitution (Albrach 1979, Corstjens 1991, pp. 213-214). As illustrated in Figures 1 and 2, price is certainly a dominant (if not the dominant) predictor in generic substitution following patent expiration in the antidepressant category. Advertising and promotion are much less important, as is demonstrated by the total suspension of advertising support for Adapin within several months of its patent expiration.
Figure 1

Extended Unit Sales Over Time for Adapin and Sinequan
Figure 2
Extended Unit Retail Prices
Over Time (25 mg)
In terms of functional form, we were confined to linear specifications. Multiplicative forms could not be used, because the logarithmic transformations intended to linearize the models and enable least squares estimation were impossible with the seasonally-adjusted residual series. With these constraints, three dynamic response specifications were considered:

**Model 1:**

\[ S_t = \alpha + \beta_1 R_{Pt} + \gamma_1 IRP_t + \delta S_{t-1} + u_t \]

**Model 2:**

\[ S_t = \alpha + \beta_2 R_{P_{t-1}} + \gamma_2 IRP_{t-1} + \delta S_{t-1} + u_t \]  

(16)

**Model 3:**

\[ S_t = \alpha + \beta_1 R_{Pt} + \beta_2 R_{P_{t-1}} + \gamma_1 IRP_t + \gamma_2 IRP_{t-1} + \delta S_{t-1} + u_t \]

where \( S_t \) denotes extended unit sales in period \( t \), \( R_{Pt} \) denotes relative price in period \( t \), and \( IRP_t \) denotes indirect relative price in period \( t \). The dynamic sales character through inclusion of lagged sales is consistent with drug response modeling in Montgomery and Silk (1972) and Parsons and Van den Abeele (1981).

The relative price variables were defined as

\[ R_{Pt} = P_t / P_{Gt} \]

and

\[ IRP_t = P_t / P_{I_t} \]

where \( P_t \) denotes the retail price for the drug in period \( t \), \( P_{Gt} \) denotes the weighted average of generic prices in period \( t \) (with weights equal to the extended unit sales), and \( P_{I_t} \) denotes the weighted average of branded drug prices in period \( t \). Hence, a distinction is made between direct competition comprised of chemically identical generic drugs, and indirect competition comprised of other branded drugs which are sold for the same indication in the therapeutic class. The direct competition covers all generic drug manufacturers who sell the generic equivalent of the branded drug (e.g., all Doxepin manufacturers for Sinequan). The indirect competition covers all antidepressant drugs listed in Table 1.
For all data series, the three models listed in (16) were estimated using least squares regression. Those for which the overall model test (F-test) was significant at 0.10 were subjected to a model selection process. The objective was to select the best model among the structurally equivalent specifications. Models 1 and 2 are nested versions of model 3; hence, a straightforward F-test for nested alternatives was applied. The test statistic equals

\[ F = \frac{(\text{ESS}_1 - \text{ESS}_3) / (k-g)}{\text{ESS}_3 / (n-(k+1))} \]

where ESS\(_i\) denotes the error sum of squares of nested model \(i\), \(g\) denotes the number of parameters in the nested model, \(k\) denotes the number of parameters in model 3 (the complete model), and \(n\) denotes the sample size. Under the null hypothesis, the test statistic has an F distribution with degrees of freedom \((k-g)\) and \([n-(k+1)]\).

If the null hypothesis was rejected, model 3 was retained as best model. If the null hypothesis could not be rejected for either nested alternative, the reduced models 1 and 2 were compared using Amemiya's Prediction Criterion (Judge et al. 1982, p. 603). This search procedure for non-nested model alternatives relies on a criterion which is based on the unconditional mean squared prediction error, or

\[ \text{APC} = \frac{(\text{TSS})}{n} \left( \frac{n+g}{n-g} \right) (1-R^2) \]

where TSS denotes the total sum of squares, and the other elements are defined as above. This criterion needs to be evaluated for both nested models, with the one providing the lowest value being retained as the best model.

The selection results are summarized in Table 2. Of the 33 estimated models (11 time series x 3 model specifications), 10 passed the significance cutoff point of 0.10. As none of the models for Ludiomil or Adapin passed this test, these two drugs were excluded from further analyses. The remaining 10 models were subjected to the selection procedure just described. For each of the four drug/sizes, model 2 was identified as the best model. Hence, both direct and indirect price competition enter the response model with a one-period lag. The least squares estimation results for those four models are contained in Table 3. As can be seen there, the lagged relative price has a significant negative impact on extended unit sales, except in the case of Norpramin 10. In this latter instance, the response parameter is positive but relatively small and insignificant. For the indirect relative price, the impact on extended unit sales is positive except in the case of Norpramin 10 where the parameter is negative and significant at 0.10. Lagged extended unit sales is generally positive but is not significant for Sinequan 10 and Sinequan 25. For Norpramin 10, lagged sales is related negatively to current
<table>
<thead>
<tr>
<th>Drug</th>
<th>Number of Observations</th>
<th>Sums of Squares</th>
<th>F-test (Nested Models)</th>
<th>Amemiya Prediction Criterion (Non-Nested Models)</th>
<th>Best Model Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model 1 (R²)</td>
<td>Model 2 (R²)</td>
<td>Model 3 (R²)</td>
<td>Model 1 Versus Model 3</td>
</tr>
<tr>
<td>Sinequan 10</td>
<td>41</td>
<td>1289447 (0.79)</td>
<td>1291283 (0.79)</td>
<td>1304532 (0.80)</td>
<td>0.205(2,35)⁩</td>
</tr>
<tr>
<td>Sinequan 25</td>
<td>41</td>
<td>8288018 (0.75)</td>
<td>8866456 (0.80)</td>
<td>8876555 (0.80)</td>
<td>1.243(2,35)</td>
</tr>
<tr>
<td>Sinequan 50</td>
<td>41</td>
<td>6739612 (0.80)</td>
<td>6846088 (0.82)</td>
<td>7011116 (0.83)</td>
<td>0.705(2,35)</td>
</tr>
<tr>
<td>Norpramin 10</td>
<td>30</td>
<td>-</td>
<td>16398 (0.22)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

---

a Only models significant (F-test) at the level 0.10 are shown.

b Degrees of freedom of F-test.
sales and in a rather significant manner. Overall, the fits are good for the Sinequan brands ($R^2$ ranging from 0.79 to 0.82) but poor for Norpramin 10 where the model is significant at 0.10 but the $R^2$ is only 0.22.

For the Sinequan brands, the results have significant face validity. The generic price substitution is captured in significant negative response parameters for the lagged relative price variable. The significant positive indirect relative price parameter captures a preferential price position of Sinequan relative to other branded drugs in the antidepressant category. The average values for the indirect relative price variable for Sinequan range from 0.73 to 0.91 with the magnitude of the price advantage increasing with strength size and, in all instances, increasing over time. This favorable position in the antidepressant category is entirely consistent with substitution captured in the positive response parameter. Interestingly enough, the advantageous price position is a post-patent-expiration phenomenon. For both Sinequan 10 and Sinequan 25, values for the indirect relative price variable are larger than one until the time that the generic equivalent showed up on the market. For Sinequan 50, the values were already less than one, indicating relatively favorable pricing among high-strength branded antidepressant drugs prior to the patent expiration date. The latter could explain the much smaller and less significant response parameter for indirect relative price for Sinequan 50.

Note that a similar argument could be used to explain the parameters for Norpramin 10. Because of its premium pricing following patent expiration, Norpramin 10 apparently lost customers to other branded antidepressants; it seemed to be relatively unaffected, however, by generics. The parameter for indirect relative price thus is negative for Norpramin 10 whereas that for direct relative price is not significant. A casual investigation of extended unit sales patterns across all drugs over time (see Figure 1 for Adapin and Sinequan) shows quite strong generic substitution for higher strength versions versus lower strength versions. This pattern is strongly borne out by Norpramin 10.

In sum, the estimated models appear intuitive. A substantive question arises, however, with respect to the transferability issue. Since Norpramin is premium priced and Sinequan is not, we clearly cannot consider these historical cases, nor their response estimates, to be identical. Even without subjecting the parameter vectors to the statistical tests suggested above, the estimates in Table 3 suggest that the parameters for Norpramin 10 and those for the Sinequan drugs do not come from the same multivariate density. This raises the question, therefore, as to which historical cases are relevant for Asendin’s post-patent response environment. Currently, Asendin is premium priced with values of the indirect relative price variable well above one.
Table 3
Model Estimation Results: Extended Unit Sales \(^a\)

<table>
<thead>
<tr>
<th></th>
<th>Sinequan 10</th>
<th>Sinequan 25</th>
<th>Sinequan 50</th>
<th>Norpramin 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-300.795</td>
<td>1993.44</td>
<td>2010.153**</td>
<td>694.527*</td>
</tr>
<tr>
<td></td>
<td>(-0.376)(^b)</td>
<td>(1.345)</td>
<td>(1.831)</td>
<td>(1.849)</td>
</tr>
<tr>
<td>Relative Price at t-1</td>
<td>-116.707*</td>
<td>-537.738*</td>
<td>-321.024*</td>
<td>3.886</td>
</tr>
<tr>
<td></td>
<td>(-2.397)</td>
<td>(-5.140)</td>
<td>(-3.692)</td>
<td>(0.495)</td>
</tr>
<tr>
<td>Indirect Relative Price at t-1</td>
<td>1648.456*</td>
<td>2241.737**</td>
<td>298.274</td>
<td>-597.891*</td>
</tr>
<tr>
<td></td>
<td>(2.159)</td>
<td>(1.661)</td>
<td>(0.205)</td>
<td>(-1.872)</td>
</tr>
<tr>
<td>Lagged Dependent Variable</td>
<td>0.127</td>
<td>0.072</td>
<td>0.386*</td>
<td>-0.389*</td>
</tr>
<tr>
<td></td>
<td>(0.831)</td>
<td>(0.518)</td>
<td>(2.655)</td>
<td>(-2.198)</td>
</tr>
<tr>
<td>R(^2)</td>
<td>0.792</td>
<td>0.801</td>
<td>0.815</td>
<td>0.215</td>
</tr>
<tr>
<td>F-Value</td>
<td>46.825*</td>
<td>49.689*</td>
<td>54.337*</td>
<td>2.371**</td>
</tr>
</tbody>
</table>

\(^a\) Model 2, post-patent expiration.

\(^b\) t-values.

* Significant at 0.05; ** Significant at 0.10.
If this price differential relative to other branded drugs is following patent expiration, then Norpramin 10 contains relevant information on sales response to indirect price competition. Norpramin 10 contains little information, however, on sales response to generic price competition. Given the complete lack of information about generic substitution in the Asendin data, therefore, inferences about generic price competition made on the basis of a transfer model using Norpramin 10 as a prior would, at best, be exceedingly weak. The additional fact that Norpramin 10 represents a single historical case which would not be sufficient for deriving $\Sigma$ in (15), led to a decision to exclude this drug from further analysis.

In selecting the Sinequan cases as historical analogies for Asendin, we obtained information about the effects of generic price competition on antidepressant sales. The question remains, however, as to whether the Sinequan cases contain relevant information about the effects of indirect price competition on sales of Asendin. As stated above, Sinequan is preferentially priced relative to other branded antidepressants whereas Asendin currently is not. However, given that historical data are available for the Indirect Relative Price of Asendin, we would expect the transferability methodology to discount the prior information from Sinequan and give more weight to the sample information from Asendin. If the pre- and post-patent-expiration price differentials for Asendin are similar, this will improve the accuracy of the transfer model. If they are dissimilar, on the other hand, the model’s accuracy will decline. With this caveat in mind, the three Sinequan time series were selected for further analysis.

Once the historical cases had been identified, the next step was to test for the assumptions underlying the transferability methodology. As discussed above, two questions arose: (a) do the parameter vectors of the historical cases come from a single multivariate distribution (i.e., the validity of expression (4)), and (b) are the parameters independent (i.e., is $\Sigma_{12}$ in (5) equal to zero). With only three historical cases, the statistical tests described above cannot be implemented because of low degrees of freedom. Even confining ourselves to the slope coefficients, we have a vector of sample means of dimension (3 x 1) and a variance-covariance matrix of dimension (3 x 3). The first test would be the homogeneity test on subsample variance-covariance matrices. Taking two historical cases at a time with replacement, we could derive three subsample variance-covariance matrices. But each of these matrices is estimated with one degree of freedom which is too few for the chi-square approximation to the standard test for equality of variance-covariance matrices (Morrison 1976, p.252).

The second test would be on the mean vector of the multivariate density. Three subsample estimates could be derived again, and under the assumption of equal but unknown variance-covariance matrices, we could rely on the Hotelling $T^2$ statistic to test for pairwise
equality as shown in Anderson (1958 p. 109). However, as each subsample estimate of the mean vector is based on only two historical cases, there are too few degrees of freedom to invoke the F approximation to enable inferences to be drawn on the pairwise equality of the subsample means.

The third test is the independence test and, as discussed above, is based on the sample variance-covariance matrix. The objective is to test whether or not $\Sigma_{12}$ in (5) is significantly different from zero. In the illustration here, $\Sigma_{12}$ is a $(2 \times 1)$ vector containing the covariances between the relative price parameter and the indirect relative price parameter, and the relative price parameter and the lagged sales parameter (i.e., in notation of (16), we have the covariance between $\beta_2$ and $\gamma_2$, and the covariance between $\beta_2$ and $\delta$). Since $\Sigma_{12}$ is a vector, we could test the hypothesis of independence with the F-test on $R^2$ (Morrison 1976, p. 108). As shown in Morrison (1976), however, we again have too few observations to be able to use the test. Thus, a statistical test of the underlying assumptions is not feasible in this illustration because we are confined to three similar historical cases. For illustrative purposes we will accept them at face value and proceed with the transfer illustration. This limitation of our data set, however, reveals one of the difficulties that arises in trying to implement the transferability methodology.

The sample estimate of the moments of the multivariate density are shown in Table 4. The transferability results are summarized in Table 5. Allowing the intercept to be free, two sets of partial transfer estimates are shown for Asendin 25 and Asendin 50. One set of estimates was derived with the individual priors (i.e., expression (8)), whereas the other was derived with the average prior (i.e., expression (9)). The corresponding estimates are not very different, indicating that the priors had been estimated with roughly equal efficiency. As discussed above, individual priors are preferred in general and we will continue the discussion of the results based on those priors.

For Asendin 25, all parameters are highly significant except the Indirect Relative Price parameter which suggests that Asendin 25 sales are not affected by the relative pricing of other branded antipressant drugs. Note that relative to the prior values shown in Table 4, the Relative Price parameter is smaller (-259.632 versus a prior value of -325.156), whereas the lagged sales parameter is larger (0.363 versus a prior value of 0.195). Hence, relative to the Sinequan cases, 1) the generic substitution for Asendin 25 is expected to be less, 2) there is neither a gain nor a significant sales loss from its pricing relative to other branded antidepressants, and 3) the carryover (or inertia) in the sales series is larger. For Asendin 50, the pattern of results is the same except that the Relative Price parameter is marginally significant (at the 0.25 level), and that the Indirect Relative Price parameter is significant but negative. The latter is different from the prior Sinequan cases (see Table 3) but is consistent
with the Norpramin 10 results; it reflects the premium pricing of Asendin relative to other branded antidepressant drugs.

The pattern of estimates is in line with expectation. As discussed above, the transferability analysis discounted the prior information on indirect price competition because of a poor match with the Asendin sample data. Unlike Sinequan, which has a price advantage over the other antidepressants, Asendin has historically been priced somewhat high. As the estimated Indirect Relative Price parameter is negative (and away from the positive prior), the negative covariances in $\Sigma$ in Table 4 result in a positive transfer in the Relative Price parameter (which becomes less negative) and the carryover parameter (which becomes more positive). Substantively, the transfer estimates for both Asendin 25 and Asendin 50 seem to suggest that we can expect significant generic substitution over time. In the case of Asendin 50, this will be compounded with a sales loss because of premium pricing versus other branded antidepressants. As there is no gain from a favorable price positioning as in the Sinequan cases, we expect accelerated sales losses beyond the Sinequan levels shown in Figure 1.

To check the validity of this suggestion, a supplemental data set was obtained from IMS containing extended unit sales and drugstore prices for Asendin as well as the prices of generic equivalents from August 1989 (the month in which Asendin's patent expired) until December 1991. The sales loss pattern for the Asendin drugs is indeed sharper than for the Sinequan cases. One year after generic introduction, the monthly extended unit sales for Asendin 25 and Asendin 50 are, respectively, 67% and 58% of the level during the last month under patent (July 1989); for Sinequan 10, 25, and 50, the respective percentages are 84, 80, and 74. Two years after generic introduction, the results for Asendin 25 and 50 are 42% and 39% with Sinequan 10, 25, and 50 at 73%, 62%, and 57%. In terms of the magnitude of sales loss after patent expiration, the Asendin drugs are closer to the Adapin drugs than to the Sinequan drugs.

Using the transferability results, the extended unit sales were predicted for the post-patent period and then were compared to actual sales. The predictions were derived using the model 2 specification with the transfer estimates shown in Table 5. For Asendin 25, the insignificant Indirect Relative Price was not included. Since the predictions are done not knowing the post-patent competitive environment, we had to decide on values to use for the Relative Price and Indirect Relative Price variables (the latter for Asendin 50 only). Indirect Relative Price was kept constant at the level just prior to the patent expiration date. For Relative Price, the focus for evaluating generic substitution, we could simulate different series reflecting the price dispersion anticipated between the Asendin drugs and their generic equivalents. For the validation here, we incorporated the actual series obtained from IMS.
Table 4
Sample Estimates (Slope Parameters)

1. **Sample Mean** $\left( \hat{\beta} \right)$

$$\hat{\beta} = \begin{pmatrix} -325.1563 \\ 1396.1556 \\ 0.1950 \end{pmatrix}$$

2. **Variance-Covariance Matrix** $(\hat{\Sigma}$ in (15))

$$\hat{\Sigma} = \begin{bmatrix} 29553.0553 & -43900.0219 & 4.2541 \\ -43900.0219 & 661335.7987 & -110.2861 \\ 4.2541 & -110.2861 & 0.0187 \end{bmatrix}$$
Table 5
Transferability Results:
Recursive Estimation for Asendin 25 and Asendin 50 \(^a\)

<table>
<thead>
<tr>
<th></th>
<th>Asendin 25</th>
<th></th>
<th>Asendin 50</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Prior</td>
<td>Individual Prior</td>
<td>Average Prior</td>
<td>Individual Prior</td>
</tr>
<tr>
<td>Intercept</td>
<td>677.632</td>
<td>627.756</td>
<td>842.176</td>
<td>840.471</td>
</tr>
<tr>
<td></td>
<td>(3.073)*b</td>
<td>(2.768)*</td>
<td>(3.987)*</td>
<td>(4.116)*</td>
</tr>
<tr>
<td>Relative Price at t-1</td>
<td>-268.900</td>
<td>-259.632</td>
<td>-214.535</td>
<td>-221.185</td>
</tr>
<tr>
<td></td>
<td>(-2.270)*</td>
<td>(-2.110)*</td>
<td>(-1.109)c</td>
<td>(-1.144)c</td>
</tr>
<tr>
<td>Indirect Relative Price at t-1</td>
<td>-75.964</td>
<td>-56.158</td>
<td>-209.912**</td>
<td>-210.453**</td>
</tr>
<tr>
<td></td>
<td>(-0.490)</td>
<td>(-0.344)</td>
<td>(-1.736)**</td>
<td>(-1.782)**</td>
</tr>
<tr>
<td>Lagged Extended Unit Sales</td>
<td>0.338</td>
<td>0.363</td>
<td>0.687</td>
<td>0.688</td>
</tr>
<tr>
<td></td>
<td>(5.754)*</td>
<td>(6.109)*</td>
<td>(11.211)*</td>
<td>(11.641)*</td>
</tr>
<tr>
<td>R²</td>
<td>0.983</td>
<td>0.980</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>F-Value</td>
<td>939.556*</td>
<td>863.707*</td>
<td>1563.191*</td>
<td>1624.029*</td>
</tr>
</tbody>
</table>

\(^a\) Estimation done with free intercept (i.e., transfer confined to slope parameters).

\(^b\) t-values; *Significant at 0.05; ** Significant at 0.10.

\(^c\) Significant at 0.25.
Two sets of predictions were derived. In one case, a continuous prediction was derived starting with the last actual sales figure prior to generic introduction, then predicting each figure into the future substituting lagged sales in the model with the predicted value of the last period. In the second case, a one-period-ahead forecast was derived following the same procedure except that actual lagged sales were substituted in. The second forecast is likely to be relatively less smooth as it adjusts with a lag to the actual series. Note, however, that the actual series are not seasonally adjusted whereas the forecasts are based on transfer estimates obtained from seasonally adjusted series. Hence, we can expect the forecasts to miss some seasonal outliers. The forecasts are shown in Figure 3.

In all cases, the sales trend is captured quite accurately. For Asendin 50, the loss is steeper than for Asendin 25, a pattern consistent with historical substitution patterns discussed above. The level of substitution is overpredicted for Asendin 25 (although actual sales were well within the 95% confidence interval of the prediction, not shown here). For Asendin 50, the forecasts are much better. Considering that the forecasts are based on seasonally adjusted series, and that the Indirect Relative Price was kept constant, the predictions are extremely good. This is especially the case for the one-period-ahead forecast.5

Despite the Sinequan drugs not being perfectly analogous to the Asendin cases as discussed above, the illustration is quite enlightening. The transfer estimates are intuitive, and for Asendin 50 they predict post-patent generic substitution and sales loss quite accurately. For Asendin 25, the trend predictions are perfect, but the level is somewhat overpredicted. The latter is probably attributable to the lack of analogy for which the transferability methodology cannot fully adjust; it illustrates the care with which historical analogous situations need to be identified. From a substantive perspective, the transfer results do indicate (and quite accurately capture, for one of the two drugs) a sales-loss pattern quite different from the historical Sinequan drugs. We have sharply decreasing monthly sales partly because of generic substitution and partly because of inferior pricing relative to other branded antidepressants. In fact, the generic substitution loss by itself is smaller than in the corresponding Sinequan cases, but in the latter we observed a significant substitution gain from other branded drugs because of advantageous pricing.

6. Conclusion

Historical analogy has been relied upon extensively in the evaluation of strategic decisions before they are implemented. Managers of a drug that will soon come off patent can analyze other drugs that have lost their patent protection in recent years. This analysis can help them evaluate the impact of generic competition that they are likely to face. To the extent that these historical cases are analogous to the current environment, the information or knowledge
Figure 3
Actual Versus Predicted Asendin Extended Unit Sales
contained therein can be transferred into the new situation. That knowledge can be strategically valuable as it allows managers to prepare for the advent of generic competition.

The methodology of data transferability discussed in this paper is an innovative application of the recursive Goldberger-Theil estimator to the general problem analysed by Aigner and Leamer (1984). The information about the historical cases is incorporated in the form of prior estimates that are updated with partial information known about the new situation. A key to the efficient estimation is the analogy of the historical cases. Econometrically, historical analogy comes down to structurally identical models and parameter dependency. Where the former can be engineered, and possibly limited to a part of the response model, the latter is strictly necessary for the methodology to work. Both can be tested with available test-statistics assuming the multivariate normal distribution captures the parameter dependence. In practice, however, it may be difficult to identify a sufficient number of historical cases to achieve the degrees of freedom necessary to carry out these tests. Moreover, efficient and valid transfer estimation requires the analogy to go beyond econometric specifications. The empirical illustration in this paper amply illustrates the importance of similarity in environments (i.e., levels and relative magnitudes of the independent predictors). To the extent that efficient individual prior cases which are analogous to the new situation can be updated with efficient sample estimates, managerially useful transfer estimates can be obtained.
FOOTNOTES

1. BLUE has to be interpreted here somewhat differently from the traditional regression model with non-random parameters. The estimates here are unbiased in the sense that $E[\hat{\beta}_i - \beta_i] = 0$ for $i = 1, 2, ..., m$ and best in the sense that the variance-covariance matrix of the prediction error of any other linear unbiased predictor exceeds the variance-covariance matrix of $E[\hat{\beta}_i - \beta_i] = 0$ by a non-negative definite matrix. Note that for simplicity, we assume that the variance-covariance matrix of the disturbances $u_i$ (for $i = 1, 2, ..., m$) is a scalar covariance matrix. This does not constrain the methodology in any way, except that for a full covariance matrix, OLS estimates would have to be replaced by GLS estimates.

2. It is not uncommon to face such tradeoffs in empirical estimation. Ridge regression, which has seen a number of useful applications in marketing (see, e.g., Mahajan, Jain, and Bergier 1977, Erickson 1981, Ofir and Khuri 1986), is based on a tradeoff between bias and efficiency.

3. Lack of sufficient observations prevented us from incorporating Surmontil as an individual historical case. Tofranil and Elavil were not considered because they are pre-Waxman Hatch Bill (1984) expirations.

4. Relying on the chi-square statistic for homogeneity of variance-covariance matrices discussed above (Morrison 1976, p.252), the pairwise values of the test statistic range from 5.88 to 79.95. With 6 degrees-of-freedom and taking into account the sensitivity to deviations from normality, the results do not conclusively reject the null hypothesis of homogeneity.

5. Seasonal outliers coincided with the identified peak periods of March-April (tax season), the summer months, and November-December.
Appendix 1: Derivation of Variance - Covariance Matrix of $U_{i0}$

From expression (6), we have

$$\beta_i - \beta_o = e_i - e_o$$

for $i = 1, 2, ..., m$.

Accordingly,

$$\beta_i = \beta_o + V_{i0}$$

for $i = 1, 2, ..., m$ (1-1)

where $V_{i0} = e_i - e_o$.

Using the extended Gauss-Markov theorem, we obtain from (1)

$$\hat{\beta}_i = (X_i'X_i)^{-1}X_i'y_i$$

for $i = 1, 2, ..., m$

which are the OLS estimates of the parameter vectors $\beta_i$ in (1).

Alternatively,

$$\hat{\beta}_i = \beta_i + (X_i'X_i)^{-1}X_i'u_i$$

for $i = 1, 2, ..., m$

which for $r_i = (X_i'X_i)^{-1}X_i'u_i$ results in

$$\hat{\beta}_i = \beta_i + r_i$$

$i = 1, 2, ..., m$ (1-2)

Substituting $\beta_i$ in expression (1-1) with an expression derived from (1-2), we obtain

$$\hat{\beta}_i = \beta_o + r_i + V_{i0}$$

for $i = 1, 2, ..., m$

or

$$\hat{\beta}_i = \beta_o + U_{i0}$$

for $i = 1, 2, ..., m$ (1-3)

where $U_{i0} = r_i + V_{i0}$.

Since $E(u_i) = E(e_i) = 0$ for all $i$, $E(U_{i0}) = 0$ 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

$$E(U_{i0}U_{i0}') = \Sigma_i^2 = 2\Sigma + \sigma_i^2(X_i'X_i)^{-1}$$

for $i = 1, 2, ..., m$

and

$$E(U_{i0}U_{j0}') = \Sigma$$

for $i \neq j$. 
Appendix 2: Comparison of Recursive Least-Squares Estimator to Aigner-Leamer Bayesian Estimator

The Bayesian estimator (given normality and quadratic loss) derived by Aigner and Learner (1984) is a posterior mean vector expressed as

$$\tilde{\beta}_o = \left[ \frac{1}{\sigma_o^2} X_o'X_o + \Sigma^{-1} \right]^{-1} \left[ \frac{1}{\sigma_o^2} X_o'y_o + \Sigma^{-1} \tilde{\beta} \right]$$

where

$$\tilde{\beta} = \left[ \sum_{i=1}^{m} \frac{1}{\sigma_i^2} X_i'X_i + \Sigma^{-1} \right]^{-1} \left[ \sum_{i=1}^{m} \frac{1}{\sigma_i^2} X_i'y_i \right].$$

Note that expression (2-1) is similar in form to the recursive estimators expressed in (10) and (11). Now, since $X_i'X_i \tilde{\beta}_i = X_i'y_i$ for $i = 1, 2, \ldots, m$, $\tilde{\beta}$ can be expressed as

$$\tilde{\beta} = \left[ \sum_{i=1}^{m} \frac{1}{\sigma_i^2} X_i'X_i + \Sigma^{-1} \right]^{-1} \left[ \sum_{i=1}^{m} \frac{1}{\sigma_i^2} X_i'y_i \right].$$

which shows that $\tilde{\beta}$ is a matrix-weighted average of the OLS estimates $\tilde{\beta}_i$ for $i = 1, 2, \ldots, m$. This estimate is different from the $\hat{\beta}$ estimate and the $\tilde{\beta}$ estimate incorporated, respectively, in the recursive estimators (10) and (11), as $\hat{\beta}$ contains the individual priors $\hat{\beta}_i$ ($i = 1, 2, \ldots, m$), and $\tilde{\beta}$ contains the midpoint of the line segment joining these individual priors (Chamberlain and Learner 1976).

The relationship between $\tilde{\beta}, \hat{\beta},$ and $\tilde{\beta}$ has been discussed by Learner (1978) for a two-dimensional problem (i.e., $m = 2$). Since the variance-covariance matrix $\Sigma$ is essentially unknown, it needs to be estimated empirically. For each estimate, the posterior mean $\tilde{\beta}$ will be different. One lemma proven by Learner (1978, p. 183) is of particular importance as it describes the area to which the posterior means will be confined. Lemma 5.1 states that all posterior means $\tilde{\beta}$ are bounded by a parallelotope (a multidimensional generalization of a parallelogram) with priors $\hat{\beta}_1$ and $\hat{\beta}_2$ at opposite vertices. Furthermore, the reverse is true. Hence, any point in the parallelotope is a posterior mean including $\hat{\beta}$, which is at the centroid of the parallelotope. Although not proven here, it seems that this lemma can be generalized to instances where $m > 2$. Moreover, the individual priors contained in $\tilde{\beta}$ are each at the boundaries of possible posterior means, where $\tilde{\beta}$ is at the centroid of those means.
REFERENCES


