THE USE OF FRONTIER ESTIMATION IN DIRECT MARKETING

Dominique Haughton
Jonathan Haughton
Alison Kelly-Hawke
Tim Moriarty

DOMINIQUE HAUGHTON is at Bentley College, Waltham, Massachusetts.

JONATHAN HAUGHTON AND ALISON KELLY-HAWKE are at Suffolk University, Boston, Massachusetts.

TIM MORIARTY is with Net-Genesis Corporation, Cambridge, Massachusetts.

Thanks are due to Paul Jones for research assistance.

Please address all correspondence to Jonathan Haughton, Department of Economics, Suffolk University, 8 Ashburton Place, Boston MA 02108. Tel: (617) 573 8750. Fax: (617) 420 4272.

ABSTRACT

A common problem in direct marketing is to identify which physicians are the best prospects for an intervention that would encourage them to prescribe a drug. The standard procedure is to measure how far their prescribing behavior falls short of the level predicted by a regression line. We suggest that a better approach is to determine how far they fall short of "best practice," as measured by a frontier line. We discuss ways of measuring the frontier and apply the techniques to both simulated data and a live data set. The results show that frontier estimation is particularly valuable when the data are heteroscedastic, a relatively common situation.

JOURNAL OF INTERACTIVE MARKETING
VOLUME 14 / NUMBER 2 / SPRING 2000

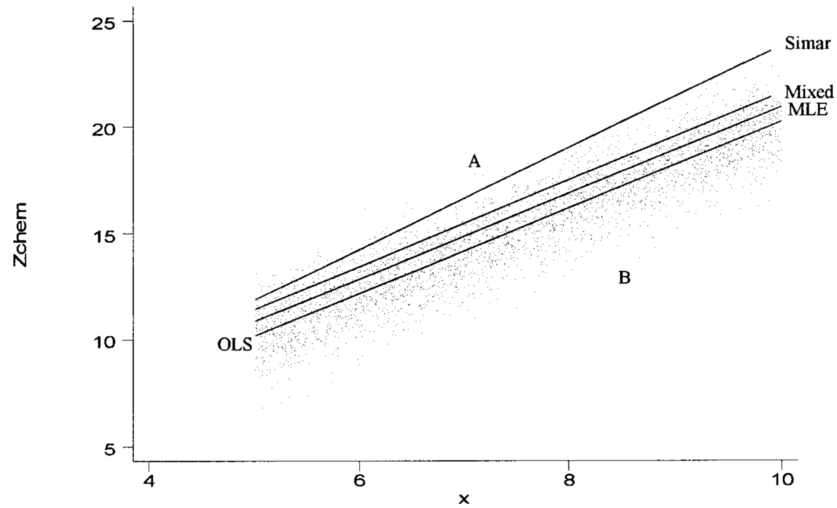

FIGURE 1
OLS and MLE lines for simulated data on drug prescriptions of Zchem as a function of variable *x*

WHY FRONTIER ESTIMATION IS NEEDED

Consider the following problem, which we encountered recently. A pharmaceutical firm sells a drug—call it Zchem—which is the only drug known to alleviate the symptoms of a common malady. Ideally the drug should be prescribed for several months after the initial diagnosis. Although almost all physicians prescribe the drug initially, only a third follow up with a substantial number of recurrent prescriptions.

The direct marketing problem is to identify the physicians that are the best prospects for an intervention, such as a mailing, which would encourage them to prescribe Zchem repeatedly. Information is available on the total amount of Zchem prescribed (in mg), the total number of prescriptions written per physician, and some other characteristics of the physician, along with the usual range of geodemographic variables.

The conventional approach is to estimate a model that has the general form

$$Y_i = f(X_i) + v_i, \qquad (1)$$

where Y_i is the i-th observation on the variable whose behavior we wish to predict, in this example the total mg of Zchem prescribed per physician per year. There are N observations (i.e., physicians) in total; X_i is the i-th observation of a $1 \times k$ vector of explanatory variables, and v_i is a random error, usually assumed to be distributed normally with mean zero and standard deviation σ. The function $f(.)$ is usually taken to be linear, and its parameters estimated by ordinary least squares (OLS). Based on the estimated equation, we can compute the predicted value of the dependent variable $\hat{Y}_i$ and compare it with the actual value, Y_i. If the predicted value is higher than the actual value, then the physician in question is prescribing less than expected. Physicians for whom $\hat{Y}_i - Y_i$ is large would be targeted for a direct mailing or other intervention.

The situation is illustrated in Figure 1, which shows simulated data for 5,000 physicians. The dependent variable, shown on the vertical axis, is mg of Zchem prescribed per physician per year. For simplicity we have assumed that the only explanatory variable is the number of older patients, X, which we suppose (arbitrarily) varies between 5 and 10. For reasons explained more fully below, the simulated data are generated according to the following:

$$Y_i = 1 + 2X_i + v_i - u_i, \qquad (2)$$