

An Alternative Reporting Style for Econometric Results*

by

Edward Leamer
UCLA

and

Herman Leonard
Harvard University

Discussion Paper 145

February 1979

*Research supported by NSF grant SOC78-09477.

An Alternative Reporting Style for Econometric Results

Estimation results which are reported in the American Economic Review and all other economics journals are selected from a large set of estimated models. Journals, through their editorial policies, engage in some selection, which in turn stimulates extensive model searching and pre-screening by prospective authors. Since this process is well-known to professional readers, the reported results are widely regarded to overstate the precision of the inferences they claim, and probably to distort them as well. Statistical analyses are, consequently, greatly discounted, and, by many, completely ignored.

This informational equilibrium is tied to the current econometric technology which generates inferences only if a precisely defined model is available, and which can be used to explore the sensitivity of inferences only to discrete changes in assumptions. The reporting of a complete sensitivity analysis is ruled out therefore first, because the econometric theory, which takes models as given, would be rendered explicitly inadequate if the sensitivity analysis were reported, and, second, because the econometric technology, if used to explore sensitivity issues, would generate vast numbers of estimated models which journals are rightfully reluctant to print.

It is the purpose of this article to discuss an alternative econometric technology, which could alter substantially the current informational equilibrium. The philosophy underlying this technology is that analyzers of nonexperimental data cannot sensibly take a model as a given. Because there are many models which could serve as a basis for a data analysis, there are many conflicting inferences which could be drawn from a given data set. If

this fact of life is acknowledged, it deflects econometric theory from the traditional task of identifying the inferences it is proper to draw from a data set given a model to the task of determining the range of inferences it is proper to draw from a data set given a range of models.

A simple introduction to this alternative econometric technology is given in Section 1 of this paper. In writing this section we have attempted to communicate the main ideas as concisely as possible. As a consequence, there is no reference to any sophisticated statistical theory and especially no mention of the Reverend Thomas Bayes. For a more complete statement as well as theological fanfare, consult Leamer (1978).

Revisionist econometric theories have been propounded from time to time, but the (unfortunate) reality is that the last econometric theorist who had a truly profound influence on the way economists report their results was Carl Freidrich Gauss. Manifold important contributions have been made since Gauss, but these have typically involved perturbations on the central theme. We believe our suggestion embodies a fundamentally distinct perspective. In an effort to make clear its value, we present three examples in Section 2. These examples are not unrepresentative of the interpretive richness which in our experience attaches to the use of the alternative technology we propose. We are prepared to rest our case primarily upon this richness as embodied in the examples. These examples were computed by a program we have named SEARCH, which is available on request.

The Statistical Theory

For pedagogical purposes, consider the linear regression model

$$Y_t = \beta x_t + \gamma_1 z_{1t} + \gamma_2 z_{2t} + u_t \quad (1)$$

where t indexes a set of T observations, u_t is assumed to be an independent normal random variable with mean zero and unknown variance σ^2 , $(Y_t, x_t, z_{1t}, z_{2t})$ is an observable vector and $(\beta, \gamma_1, \gamma_2)$ is an unobservable parameter vector. Inferences are to be drawn from a data set about the effect of the variable x on the dependent variable Y . In an ideal experiment, the variables z_1 and z_2 would have been controlled at some constant level. As a substitute for experimental control, the variables z_1 and z_2 are included in the equation.

A researcher, wishing to show that β is large or finding it difficult to estimate γ_1 and γ_2 accurately might estimate the four different regressions using different subsets of the "control" variables (z_1, z_2) , and select for reporting purposes the most favorable result. The alternative procedure, which is advocated here, is, first, to enlarge the search, and, second, to require reporting of both the most favorable and the least favorable outcomes. The search may be enlarged by defining a composite control variable

$$w_t(\theta) = z_{1t} + \theta z_{2t}$$

where θ is a number to be selected by the researcher. The regression model is now

$$Y_t = \beta x_t + \eta w_t(\theta) + u_t \quad (2)$$

Each value of θ selects a different constraint of the form $\gamma_2 = \theta \gamma_1$ and consequently a different method for estimating β . Allowing θ to take any

value contrasts with the usual search procedure in which θ is implicitly permitted to take one of only four values:

(1) if only z_1 is in the equation, then $\theta = 0$.

(2) if only z_2 is in the equation, then $\theta = \infty$.

(3) if both z_1 and z_2 are in the equation, the coefficients η and $\eta\theta$ on z_1 and z_2 must be the least squares estimates of $\hat{\gamma}_1$ and $\hat{\gamma}_2$. Using the notation that the least squares estimated coefficient of q as an explanatory influence of p in a model also containing r is $b_{pq \cdot r}$, this implies that

$$\theta = \hat{\gamma}_2 / \hat{\gamma}_1 = \frac{b_{yz_2 \cdot xz_1}}{b_{yz_1 \cdot xz_2}}$$

(4) if neither z_1 nor z_2 is in the equation then the least squares estimate of η must be 0. Denoting by M the idempotent matrix $I - x(x'x)^{-1}x'$, this implies

$$0 = \hat{\eta} = b_{yw \cdot x} = \frac{y'Mw}{w'Mw}$$

so that $0 = y'Mw = y'M[z_1 + \theta z_2]$.

$$\text{Thus } \theta = -\frac{y \cdot Mz_1}{y \cdot Mz_2} = \frac{-y \cdot Mz_1 / y \cdot My}{y \cdot Mz_2 / y \cdot My} = \frac{-b_{z_1 y \cdot x}}{b_{z_2 y \cdot x}}$$

Constraining θ to be one of $0, \infty, \frac{b_{yz_2 \cdot xz_1}}{b_{yz_1 \cdot xz_2}}$ and $\frac{-b_{z_1 y \cdot x}}{b_{z_2 y \cdot x}}$ has the virtue of

historical acceptance and the additional merit that it is comparatively easy to carry out computationally in the context of the existing econometric technology. It has no other obvious intrinsic virtues. Consequently, we now expand the search to include all values of θ .

To each value of θ there is a least squares estimate of β , $\hat{\beta}(\theta)$.

The most "favorable" value of θ , for the researcher who wishes to show

β is large, is found by maximizing $\hat{\beta}(\theta)$ with respect to θ , and the least favorable value is found by minimizing $\hat{\beta}(\theta)$. These extreme values, $\hat{\beta}_{\min}$ and $\hat{\beta}_{\max}$, delineate the ambiguity in the inferences about β induced by the uncertainty about the model. If the interval $[\hat{\beta}_{\min}, \hat{\beta}_{\max}]$ is short in comparison to the sampling uncertainty¹, the ambiguity in the model may be considered irrelevant since all models lead to essentially the same inferences. But if the bound is wide, an effort should be made to narrow the family of models, and, hopefully, to sharpen the inferences. One way to narrow the family of models is to constrain the parameters, γ_1 and γ_2 , to lie within the ellipse

$$\gamma_1^2 a^2 + \gamma_2^2 \leq r^2 \quad (3)$$

where a is the relative lengths of the two principal axes and r is the radius. This may seem initially to be a peculiar thing to do, but this constraint is the foundation of the voluminous literature on "biased estimation."² It can be justified in the following way. The only compelling reason for the omission of the z -variables is that they are thought to be doubtful. If they truly don't belong in the equation, then a better estimate of β can be produced by an equation with the z -variables omitted. To say that the z -variables are doubtful is to say that the parameters, γ_1 and γ_2 , are small. One precise definition of smallness is given by equation (3), and a natural way to estimate the parameter β is to use least-squares subject to the constraint (3). Henceforth, this constraint will be called the prior

¹An alternative definition of shortness derives from a decision problem based on β : the interval is short if all values in the interval lead to essentially the same decision.

²This includes ridge regression, minimum mean-squared-error regression and "Stein" regression.

ellipse in reference to the fact that it represents information about γ_1 and γ_2 which is available prior to the data analysis.

If the parameters of the prior ellipse, a^2 and r^2 , were known, this procedure would generate a unique estimate, but we are unaware of any real data analysis situations in which the values a^2 and r^2 could be sensibly taken as given. For any value of a and r , there is a constrained least-squares estimate, $\hat{\beta}(a, r)$, computed by minimizing the sum of squares subject to the constraint (3). We now turn to an examination of the function $\hat{\beta}(a, r)$.

Consider first the case when a^2 is known, taken without loss of generality to be equal to one. For every value of r^2 , equation (3) defines a circle located at the origin, depicted in Figure 1. Also depicted in Figure 1 are the unconstrained least-squares estimates of γ_1 and γ_2 , and the contours of equal residual sums-of-squares around $(\hat{\gamma}_1, \hat{\gamma}_2)$. It should be noted that the sum-of-squared residuals depends on the estimate of β as well as on the estimates of γ_1 and γ_2 . The contours in Figure 1 make use of the conditional least-squares estimate of β , given γ_1 and γ_2 , so that the residual sum-of-squares can be written as a function of γ_1 and γ_2 only.

The estimation problem of minimizing the residual sum-of-squares subject to the constraint (3) can be defined graphically in terms of a tangency point between a sum-of-squares ellipse and the given circle located at the origin. As the radius of the circle is varied, a curve of estimates is formed which we call an information contract curve.^{3/} This language is selected to suggest the Edgeworth-Bowley analysis of trade between a pair of consumers, a setting which is analogous to our own problem both mathematically

^{3/}This curve has been called the "ridge trace" by Hoerl and Kennard (1970) and the "curve décolletage" by Dickey (1975).

and substantively. In the Edgeworth-Bowley analysis, a contract curve represents the Pareto-efficient allocation of commodities to a pair of consumers with conflicting desires. Here, the information contract curve represents the "Pareto-efficient" set of estimates given two conflicting sources of information.

The choice of a point on the contract curve in the Edgeworth-Bowley analysis requires cardinal utility and a social welfare criterion. To put it differently, there has to be a way of comparing the utilities of the two consumers. Analogously, the choice of a point on the information contract curve requires us to compare the strength or precision of the two information sources, a problem to which we return below.

Next consider the case when neither a^2 nor r^2 can be taken as known. For any a there will be a contract curve, two of which are depicted in Figure 2. The hull of all such curves is just the shaded area, which has been shown by Leamer and Chamberlain (1976) to be a subset of the set of all weighted averages of the four regressions formed by omitting (or not omitting) the two z variables. This brings us back to the procedure which was first mentioned in this section. Now we have a justification for it: if, in the researcher's opinion (and his readers!), he thinks γ_1 and γ_2 are small in the sense of the ellipse (3), but he knows neither a^2 nor r^2 , then the extreme estimates that can be generated from the sample are the four regressions formed by omitting the z variables. We would then recommend, in fact require, that he report both the minimum and the maximum estimate of β from among this set of four regressions, and in addition we would suggest reporting the other two as well.

The widest bound for β swept out by the parameter θ can also be depicted graphically. The prior ellipses so far considered all have axes in the coordinate directions. The quadratic form (3) which determines these ellipses can be minimized by reducing γ_1 and γ_2 independently. If the quadratic form were, more generally,

$$\gamma_1^2 a^2 + \gamma_2^2 + c\gamma_1\gamma_2 \leq r^2$$

then instead of reducing γ_1 to zero, it is preferable to translate γ_1 to $-c\gamma_2/2a^2$. This quadratic form determines a tilted ellipse, which can generate estimates outside the shaded area in Figure 2. The hull of all contract curves, with all families of prior ellipses, is the shaded area in Figure 3. The boundary of this region, which is an ellipse, is the set of constrained least-squares points subject to constraints of the form $\gamma_2 = \theta\gamma_1$, where θ varies from $-\infty$ to ∞ .^{4/}

We have now discussed three bounds for the estimates of β which can be generated from a given data set. The choice among these bounds depends on how precisely the researcher is willing to define the vague notion that the z-variables are doubtful. If nothing more can be said, the widest set of estimates, depicted in Figure 3 applies. If it can be agreed that the definition of doubtful should be restricted to mean that $\gamma_1^2 a^2 + \gamma_2^2$ is likely to be small, where a^2 is left undetermined, then the shaded area in Figure 2 is the bound. If, thirdly, "doubtful" is even more precisely defined to mean that $\gamma_1^2 + \gamma_2^2$ is small, then the contract curve in Figure 1 defines the bound.

^{4/}For a proof see Leamer [1978, p. 128].

To narrow the bounds further, it will be necessary to devise a method for choosing points from a contract curve. As in the Edgeworth-Bowley analysis, this will require us to compare the strengths of the sources of information. If the sample information is relatively precise, it will be better to select points relatively close to the least-squares point. Conversely, if the prior information is relatively precise, we will prefer points in the neighborhood of the prior point, the origin in our example. One way to make the prior information comparable to the sample information is to act as if the prior information came from a previous set of observations. Suppose we observed the process

$$y_t^* = \gamma_1 z_{1t}^* + \gamma_2 z_{2t}^* + u_t^*$$

with u_t^* normally distributed with variance σ^{*2} . Suppose also that

$$\sum_t z_{1t}^* z_{2t}^* = 0, \quad \sum_t z_{it}^{*2} = 1, \quad \text{and} \quad \sum_t z_{it}^* Y_t^* = 0.$$

Then the least-squares estimate of γ_1 and γ_2 would be $(\hat{\gamma}_1^*, \hat{\gamma}_2^*) = (0, 0)$ with variance matrix $\sigma^{*2} I$.

Now consider pooling this prior sample with the sample generated by equation (1). The pooled estimates may be obtained by stacking the two samples:

$$\begin{pmatrix} \hat{\beta}^{**} \\ \hat{\gamma}_1^{**} \\ 1 \\ \hat{\gamma}_2^{**} \end{pmatrix} = \begin{pmatrix} \sigma_1^{*-2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \sigma^{-2} \begin{pmatrix} x'x & z_1'x & z_2'x \\ z_1'x & z_1'z_1 & z_1'z_2 \\ z_2'x & z_2'z_1 & z_2'z_2 \end{pmatrix} \end{pmatrix}^{-1} \begin{pmatrix} x'Y \\ z_1'Y \\ z_2'Y \end{pmatrix}$$

Using the partitioned inverse rule we can obtain

$$\begin{pmatrix} \hat{\gamma}_1^{**} \\ \hat{\gamma}_2^{**} \end{pmatrix} = \left[\sigma_1^{*-2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sigma^{-2} \begin{pmatrix} z_1' M z_1 & z_1' M z_2 \\ z_2' M z_1 & z_2' M z_2 \end{pmatrix} \right]^{-1} \begin{pmatrix} z_1' M Y \\ z_2' M Y \end{pmatrix} \sigma^{-2}$$

where $M = I - x(x'x)^{-1}x'$. This equation describes the pooled estimates of γ_1 and γ_2 as a function of the variance ratio σ^{*2}/σ^2 . If this variance ratio is small, the estimates will be close to the prior estimates (0, 0), and if this variance ratio is large, the estimates will be close to the least-squares estimates

$$\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \begin{pmatrix} z_1' M z_1 & z_1' M z_2 \\ z_2' M z_1 & z_2' M z_2 \end{pmatrix}^{-1} \begin{pmatrix} z_1' M Y \\ z_2' M Y \end{pmatrix}.$$

Moreover, as the variance ratio is varied from zero to infinity, the estimates $(\hat{\gamma}_1^{**}, \hat{\gamma}_2^{**})$ will sweep out exactly the contract curve depicted in Figure 1.

To the extent that the prior information can be considered as coming from a hypothetical normal experiment, and to the extent that we can select the variance ratio σ^{*2}/σ^2 , we are now able to pick a particular point from the contract curve.

The variance σ^2 can be estimated from the data Y, x, z_1, z_2 . The prior variance σ^{*2} presents greater difficulty. This number determines the size

of the prior confidence intervals for γ_1 since a 95% interval, for example, is $|\gamma_1| \leq 1.96\sigma^*$. In selecting a value of σ^* , it is therefore necessary for the researcher to ask himself how confident he is that γ_1 is small. If he feels very confident that $|\gamma_1| \leq 1.96$, then $\sigma^* = 1$ may be a useful starting point.

It seems to us unlikely that a precise number for σ^* could ever be selected. We recommend a sensitivity analysis in which several different values of σ^* are selected. A researcher might sensibly constrain σ^* to an interval such as $.2 \leq \sigma^* \leq 4$. As σ^* is varied in this interval, a subset of points on the contract curve is selected. Often this subset of points will be so narrow that the ambiguity that remains will be for practical purposes irrelevant.

There is another approach that can be used to narrow the bounds. It may be argued that the fictitious prior sample should not be treated on the same footing as the actual sample. We would rightfully be suspicious of prior information which implied an estimate outside the 95% sample confidence ellipsoid. This suggests repeating the preceding analysis with the additional constraint that the estimates fall within the $\alpha\%$ sample confidence ellipsoid.

To see the implications of this constraint, examine Figure 4. Given particular values of γ_1 and γ_2 , an estimate of β may be computed by regressing $y - z_1\gamma_1 - z_2\gamma_2$ on x : $\hat{\beta}(\gamma_1, \gamma_2) = (x'x)^{-1}x'[y - z_1\gamma_1 - z_2\gamma_2]$, which is a linear function of γ_1 and γ_2 . In the following discussion we assume that $x'z_2 = 0$ and consequently extreme values for $\hat{\beta}$ occur when $\hat{\gamma}_1$ is extreme. If it is desired to discuss the evidence about β , there are three points on the 95% ellipse which can be mentioned. Point C is on the contract curve. It is

the point within the 95% confidence ellipse that is most favored by the fictitious prior sample. Corresponding to point C is a unique value of σ^* which would imply point C as a pooled estimate. If this number is very small it may be inferred that point C is an unlikely pooled estimate, since the prior sample would have had to be quite informative to produce it. The two other points in the figure, A and B, represent the extreme values of $\hat{\beta}(\gamma_1, \gamma_2)$ within the 95% sample ellipse, and also within the ellipse of constrained estimates. The extreme points on the ellipse of constrained estimates, A' and B', may be very unlikely from the point of view of the sample. The points A and B, which may require an unlikely prior sample, are nonetheless reasonably acceptable from the standpoint of the data. In the fortuitious event that A, B and C are all close, it is not necessary to consider further the question of whether the priors that imply points A and B are sensible.

All of this discussion may now be summarized in a single graph, Figure 5. The horizontal scale is the probability value attaching to a given sample confidence ellipsoid. When that confidence level is zero, the sample ellipsoid is a single point; as the confidence grows, the ellipsoid grows to contain the whole space. The points A, B and C from Figure 4 can be found on the graph, with left-hand side vertical scale. The value of σ^* necessary to determine point C can be seen to be .25 from the right-hand scale. Also depicted is the pooled t-statistic of the coefficient, called the posterior t. If σ^* is equal to .25 the posterior t for β is seen to be ~~-2.~~ .8

We note in passing three simple generalizations that greatly increase the applicability of the preceding discussion. First, there is nothing special about the choice of zero as a point of departure for coefficient values. We have so far treated "doubt" about coefficients implicitly as doubt that they are different from zero, but we could as easily treat doubt that they are different from any other selected values.

The second generalization is that prior information need not apply only to individual coefficients, but may apply as well to linear combinations. This amounts to the observation that the parameterization of the model has to be selected by the researcher. For example, equation (1) can be rewritten as $Y_t = \beta(x_t + z_{1t} + z_{2t}) + (\gamma_1 - \beta)z_{1t} + (\gamma_2 - \beta)z_{2t} + u_t$, in which case, when the variable z_1 is omitted, the restriction $\gamma_1 = \beta$ is imposed. A member of the continuous set of restrictions indexed as above by θ then takes the form $(\gamma_2 - \beta) = \theta(\gamma_1 - \beta)$.

The third generalization is that we need not focus on a particular coefficient value as the issue of interest; we can as easily treat the inferential ambiguity pertaining to any linear combination of coefficients. This, like the second generalization, can be expressed merely as the problem of choosing the parameterization.

Examples

In this section, we offer three examples of the alternative method and reporting style which we recommend. The numbers reported are all computable by a regression package which we have named SEARCH (Seeking Extreme and Average Regression Coefficient Hypotheses).

Doubtful Variables^{5/}

It is very common to have a model with a few explanatory variables that are known to belong in the equation and a longer list of "doubtful" explanatory variables. The first set of variables is likely to be the focus of the analysis, and the second set is used to "control" for other influences. If the list of doubtful variables is long, estimation with all the doubtful variables included in the equation produces large standard errors on the coefficient of the "focus" variables. In this situation, it is typical to try different

^{5/}This material is drawn from Leamer (1978, p. 194-197).

subsets of the doubtful variables, and it is hoped that the coefficients of the focus variables will not change much as the list of doubtful variables is changed. But this search is both haphazard and nonexhaustive. Furthermore, if the coefficients of the focus variables change very much, this ad hoc search does not suggest how to average the many computed estimates into a single number.

SEARCH is ideally suited to deal with this problem. The bounds that the program reports are the extreme estimates of the focus coefficients with ideally chosen doubtful variables included in the equation. There is no way of "fiddling" with the doubtful variables to get an estimate outside the reported range. The points on the contract curve reported by the program are mixtures of the 2^q regressions that could be computed using subsets of the q doubtful variables. Thus the program searches exhaustively the set of possible regressions and also suggests weighted averages of the regressions, the latter being important when the bounds are wide.

The following example has eight "doubtful" regional dummy variables. The dependent variable is the wage rate, and the focus variables are the education of the wage earner, his age, and the square of his age. A dummy variable for a region is necessary if the labor market in the given region is "separated" from the markets in other regions. To say that the dummy variables are doubtful is to say that in the absence of evidence to the contrary, we should view the labor market as a national market.

The estimated model with all the dummy variables included is (standard errors in parentheses):

$$\begin{aligned}
 W = & .041D_1 + .098D_2 + .051D_3 & & -.019D_4 \\
 & (.34) & (.32) & (.46) & & (.34) \\
 & +.004D_5 & -.178D_6 & +.086D_7 & & +.060D_8 \\
 & (.46) & (.43) & (.50) & & (.35) \\
 & +.05EDUC + .137 AGE - .0015(AGE)^2 + 5.737 \\
 & (.30) & (.047) & (.0006) & & (.96)
 \end{aligned}$$

where

- D_1 = Mid-Atlantic
- D_2 = East North Central
- D_3 = West North Central
- D_4 = South Atlantic
- D_5 = East South Central
- D_6 = West South Central
- D_7 = Mountain
- D_8 = Pacific

(New England omitted)

The bounds for the coefficients of the three focus variables are reported in the table below. The numbers in parentheses are the standard errors of these coefficients if the model that implied the estimate could be taken as given. (Remember that these bounds include regressions subject to constraints such as $\beta_1 = \beta_2$, which says the Mid-Atlantic and East North Central regions can be aggregated. They also include constraints of the form $\beta_1 = 0$.)

Table 1

Bounds for the Focus Coefficients using Any Combinations of Regional Dummies

	EDUC	AGE	(AGE) ²
Maximum	.0577(.0177)	.139(.029)	-.00147(.00035)
Minimum	.0446(.0178)	.131(.029)	-.00155(.00035)

For this particular problem the ambiguity in the specification does not translate into substantial ambiguity in the focus coefficients. The interval of estimates for the education coefficient is .0446 to .0577. But the sampling standard error of this coefficient in the unconstrained model is .03, which is large compared to the specification range .0577 - .0446 = .0131. To put it briefly, the sampling error is more important than the specification error.

Functional Form

The descriptive perspective of the "doubtful variables" problem just treated can readily be generalized to encompass wider problems of the functional form of the underlying relationship. An investigator may wish to see how sensitive his inferences are to changes in functional form. A common approach to this problem is to add second order and interaction terms in variables to see how much difference they make. SEARCH provides a more complete and definitive approach.

As an example, suppose we were interested in estimating the value of adding a room to a house. We might consider the log linear hedonic price index for housing attributes estimated using ordinary least squares:

$$\begin{aligned}
 \ln p_i = & .0024yr_i + .045 Modkit_i + .046 Garages_i + .049 Hearths_i \\
 & (.0002) \quad (.009) \quad (.006) \quad (.006) \\
 & +.0029 Lot_i + .24 Floor_i + .11 Baths_i - .033 Consqual_i \\
 & (.0002) \quad (.02) \quad (.02) \quad (.003) \\
 & +.19 \ln (Rooms)_i + .021 Ppuptrct_i + .15 \ln(Inc.)_i \\
 & (.03) \quad (.004) \quad (.03) \\
 & -.0015 Dist_i - .014 Floor_i^2 + .0083 Floor_i Baths_i \\
 & (.0002) \quad (.005) \quad (.008) \\
 & -.0092 Floor_i Rooms_i \\
 & (.003)
 \end{aligned}$$

where p_i is the price of the i^{th} house

yr_i is the year in which the house was built

$Modkit_i$ is 1 if house i has a modern kitchen

$Garages_i$ is the number of garages i

$Hearths_i$ is the number of fireplaces

Lot_i is the size of the lot (in 1000s of square feet)

$Floor_i$ is the floorspace (in 1000s of square feet)

$Baths_i$ is the number of baths

$Consqual_i$ is a construction quality index from 1 to 9 (with 1 best)

$Rooms_i$ is the number of rooms

$Ppuptrct_i$ is the per pupil school expenditure in the census tract m

Inc_i is the average income in the census tract

$Dist_i$ is an index of distance to employment for the town in which the house is located

i is from 1 to 2195 single family houses sold in one of thirteen suburban towns in the Boston SMSA in 197 .

The focus variables in this equation are baths, rooms, and floorspace; the issue at hand is the value of additional rooms. The last three terms ($floor^2$, $floor * baths$, and $floor * rooms$) are included because concern has often been expressed by researchers using this data that the relation between house value and these variables may be nonlinear and may involve interactions. These variables are to be treated as "doubtful"; thus, the issue to be investigated is whether it is material (in terms of inferences concerning floorspace or room value) whether the relationship between the log of house

price and floorspace, rooms, and baths is presumed to be simple and linear or, rather, nonlinear and interactive.

Since the index is log-linear, the derivative of the dependent variable with respect to an explanatory influence is the estimated percentage change in house value for a one unit change. Because the relationship is nonlinear, we must choose a point at which to describe the derivative. We focus initially on the "mean attribute house," which has the mean floorspace (1491 sq. ft.) number of rooms (6.8), and number of baths (1.6).

Table 2 shows the extreme bounds for estimates of the value of 500 sq. ft. of additional floorspace and the value of an additional room, both evaluated for the mean attribute house, under the presumption that the three quadratic terms in the equation are doubtful. Consideration of potential specification error in this form clearly generates substantial ambiguity about both issues. We are therefore forced to consider additional restrictions that may help us to limit our uncertainty.

Table 2

	Value of 500 sq. ft. of Floorspace	Value of Additional Room
Extreme Upper Bound	7.4 (.5)	1.9(.3)
Extreme Lower Bound	4.8 (.4)	1.2(.3)

We consider first the value of additional floorspace by itself. Treating the three quadratic terms in the equation as doubtful, the extreme estimates for the value of 500 square feet of floorspace (with no additional room) are 4.8% and 7.4% of the value of the house. This range of 2.6%, compared with the standard errors of about .4%, is quite wide.

The summary diagram, Figure 6, shows the constrained extreme bounds for all values of the data confidence level. We note immediately that practically all of the ambiguity (in terms of the width of the unconstrained bounds) occurs at points in the feasible ellipsoid that are highly unlikely in the view of the data. If we agree to limit consideration to those points that are both within the ellipsoid of constrained estimates and also within the 90% data confidence ellipsoid, then the range of estimates is only from 6.6% to 7.4%.

Narrowing the estimate further would require the imposition of additional information. For example, it might be believed that the statement that the quadratic terms are not important takes the special form that $\sum_{j=13}^{15} \beta_j^2$ is small. This narrows the focus from all possible contract curves for this data with these three terms doubtful to one special contract curve. If we consider the whole range of this contract curve, the range of estimates is from 4.9% to 7.3% of house value, which is almost as wide as the range given by the extreme bounds. If we restrict attention to the part of this contract curve that lies within the 90% data confidence ellipse, the range of estimates is from 6.7% to 7.3%. This range is nearly as wide as the extreme bounds within the 90% confidence ellipsoid.

Figure 6 thus shows that imposition of this special form of doubt that the three quadratic terms belong in the equation hardly reduces the inferential ambiguity at all. This is because the particular contract curve we have chosen starts almost as high as the extreme upper bound and runs very close to the extreme lower bound. This special restriction on the form of doubt is much less helpful in this instance than is the restriction of attention to estimates within the 90% data confidence ellipsoid.

We focus next on the value of an additional room (with no additional floorspace) for the mean attribute house. Here the extreme bounds are 1.2% and 1.9% of house value, with standard errors approximately .3%. This range seems smaller only because the estimates are smaller absolutely. The ambiguity due to uncertain knowledge of functional form is of the same order as the sampling uncertainty. Figure 7 shows the constrained extreme bounds over the range of data confidence ellipsoids. We note that in this case the constrained bounds increase smoothly, and that only fairly severe restrictions in terms of data likelihood provide material reductions in inferential uncertainty. For example, even by restricting attention to estimates within the 75% data likelihood ellipsoid we are left with a range of estimates from 1.3% to 1.7%, which offers hardly any improvement over the unconstrained bounds.

In this case, however, as indicated by Figure 7, the additional restriction that $\sum_{j=13}^{15} \beta_j^2$ is small limits the range to from 1.5% to 1.6%. Moreover, restricting attention to those estimates on the contract curve within the 99.9% data confidence ellipsoid reduces the range to from 1.48% to 1.53%. We conclude that the value of an additional room can be estimated with little specification ambiguity, but only if we accept a special restriction on the form of "doubt" that the estimating equation is nonlinear.

We now consider the combined effect of an additional room and 300 square feet of floorspace. To make the nonlinearities in the equation more important, we shift focus to a house larger than the mean, with 2,000 square feet of floorspace, 2 1/2 baths, and 8 rooms. The extreme estimates with no limiting assumption about the form of doubt that the equation

involves second order interactions, are 3.8% and 5.0% of house value, with standard errors of about .3%. This range is appreciable, but appears less important than that of either floor value or room value alone.

If we did wish to narrow the range further we would turn to the constrained bounds. Figure 8 shows the constrained bounds and contract curve. As in the case of value of a room alone, the constrained bounds increase smoothly as we consider less likely data confidence ellipsoids. There is again little to be gained except by imposing quite severe restrictions on the level of data confidence. As the figure indicates, however, we can reduce the ambiguity greatly by limiting consideration to the single contract curve corresponding to the restrictions that $\sum_{j=13}^{15} \beta_j^2$ is small. This limits the range to only 4.3% to 4.5%. It is interesting to note that restricting consideration to the part of the contract curve within reasonable data confidence ellipsoids is not particularly helpful, since the values from 4.3% to 4.45% are encountered within the 50% data confidence ellipsoid.

Distributed Lag Estimation

Another common problem in economics is the estimation of distributed lag processes. Consider the import demand function estimated by ordinary least squares adjusted for autocorrelation equal to .98:

$$\begin{aligned}
 M_t = & .20Y_t + 1.9Y_{t-1} - .91Y_{t-2} \\
 & (.43) \quad (.51) \quad (.52) \\
 & + .53Y_{t-3} - .34Y_{t-4} - .45P_t - .47_{t-1} \\
 & (.54) \quad (.41) \quad (.55) \quad (.58) \\
 & + .45P_{t-2} - .62P_{t-3} + .40P_{t-4} - 3.6 \\
 & (.52) \quad (.54) \quad (.51)
 \end{aligned}$$

where standard errors are in parenthesis and where

M_t = logarithm (United States imports in the t^{th} quarter divided by a price index of imports)

Y_t = logarithm (United States GNP in quarter t divided by the GNP price index)

P_t = logarithm (import price index divided by GNP price index)

t = 1951 first quarter to 1967 fourth quarter

Economists would generally expect to see the coefficients on the income variables positive and the coefficients on the price variables negative. The peculiar saw-tooth pattern of coefficients would be regarded as highly unlikely, and some constraint on the coefficients would undoubtedly be used to "improve" or to smooth the estimates. One possibility is to constrain the coefficients of each of the distributed lag patterns to lie on a line. The resulting estimates are

$$\begin{aligned}
M_t = & \alpha + .86Y_t + .57Y_{t-1} + .28Y_{t-2} - .01Y_{t-3} \\
& -.30Y_{t-4} - .54P_t - .33P_{t-1} - .12P_{t-2} \\
& + .09_{t-3} + .31P_{t-4}.
\end{aligned} \tag{3}$$

Although this constraint does eliminate the wild pattern of coefficients, it does not produce coefficients that are all the same sign for each variable. We could constrain them all to be equal, yielding the estimated equation

$$M_t = \alpha + .26 \sum_{\tau=0}^4 Y_{t-\tau} - .22 \sum_{\tau=0}^4 P_{t-\tau}. \tag{4}$$

Each of these three estimated equations is appropriate for one extreme form of information about the coefficients. Since we believe none of these three forms of information completely, we might informally mix together the three results.

For example, if it were desired to estimate the long run income and price elasticities, that is, the sum of the coefficients, it might be noted that the three estimates of the income elasticity are 1.37, 1.41, and 1.30, which are not especially different. The estimates of the price elasticity are somewhat more dispersed: -.68, -.58, -1.1. This contrasts nonetheless with individual coefficients which vary greatly from equation to equation. The three estimates of the coefficient on Y_{t-2} , for example, are -.91, .28, and .26.

These results, however, are hardly definitive. We would like to consider more general (and sensible) restrictions on the form of the lags to see if the important conclusions we wish to make are sensitive to knowledge about

the lag structure. The natural family of simple restrictions on the coefficients is $R = (\beta_1 = \beta_2, \beta_2 = \beta_3, \beta_3 = \beta_4, \beta_4 = \beta_5, \beta_6 = \beta_7, \beta_7 = \beta_8, \beta_8 = \beta_9, \beta_9 = \beta_{10})$. If all of these restrictions are imposed, the estimated equation is as reported above, Equation (3). Alternatively, these restrictions can be subtracted from each other to form the set

$[(\beta_1 - \beta_2) - (\beta_2 - \beta_3) = 0, (\beta_2 - \beta_3) - (\beta_3 - \beta_4) = 0, (\beta_3 - \beta_4) - (\beta_4 - \beta_5) = 0, (\beta_6 - \beta_7) - (\beta_7 - \beta_8) = 0, (\beta_7 - \beta_8) - (\beta_8 - \beta_9) = 0, (\beta_8 - \beta_9) - (\beta_9 - \beta_{10}) = 0]$. If these restrictions are imposed, Equation (3) is the result. If this differencing is done repeatedly, the coefficients can be restricted to lie on a polynomial of arbitrary degree. Another way of imposing smoothness on the coefficients, which captures partly the notion that the right tail of the distribution is likely to be smoother than the first coefficients, is the restriction $(\beta_i - \beta_{i+1}) = \lambda(\beta_{i+1} - \beta_{i+2}), 0 < \lambda < 1$.

Any estimates that would result from any of these restrictions are within the bounds reported in Table 3. These bounds are the extreme estimates of the indicated coefficients or indicated linear combinations, computed using linear combinations of the set of constraints R , identified in the paragraph above. Only the long-run income elasticity, $\sum_1^5 \beta_i$, turns out to be insensitive to the form of smoothness imposed on the coefficient. There is no smoothness notion which would produce an estimate for $\sum_1^5 \beta_i$ exceeding 1.49 or falling short of 1.18. The long-run price elasticity $\sum_6^{10} \beta_i$ is also relatively insensitive to the form of smoothness prior. The ranges for individual coefficients are, however, quite wide.

These bounds can be narrowed in two ways. First, a single "contract" curve of estimates may be produced by making the special assumption of Shiller (1973), that $\sum_1^4 (\beta_i - \beta_{i+1})^2 + \sum_6^9 (\beta_i - \beta_{i+1})^2$ is "small." Secondly, we may

Table 3

Bounds for Estimates of Distributed Lag Process

Using Linear Combinations of the Constraints

$$R = [\beta_1 = \beta_2, \beta_2 = \beta_3, \beta_3 = \beta_4, \beta_4 = \beta_5, \beta_6 = \beta_7, \beta_7 = \beta_8, \beta_8 = \beta_9, \beta_9 = \beta_{10}]$$

	β_1	β_2	β_3	β_4	β_5	$\sum_1^5 \beta_i$	β_6	β_7	β_8	β_9	β_{10}	$\sum_6^{10} \beta_i$
Max	1.18	2.22	.82	1.59	.86	1.49	.86	.92	1.25	.75	1.19	-.25
Min	-.72	-.06	-1.48	-.80	-.95	1.18	-1.53	-1.62	-1.02	-1.59	-1.01	-1.54

restrict our choice of estimates to lie within a sample confidence ellipsoid with a certain "acceptable" confidence value, say .95.

The contract curve is reported in Table 4. The first column of this table is the classical probability value attaching to the ellipsoid on which the indicated estimate lies. The least-squares point is the first row, and it has by definition a confidence value of zero. The last row is the constrained least squares point, indicated above by Equation (3). This point lies on the boundary of the .9696 classical confidence ellipsoid, which is perhaps farther from the unconstrained least-squares point than we would want to go. If we restrict ourselves to the 95 percent ellipsoid, the coefficients are "properly" signed and "pleasantly" smooth. Troubles, in that sense, begin when we get interior to the 75 percent ellipsoid.

The column SIGMA1 indicates the prior standard error of $\beta_i - \beta_{i+1}$ which would be necessary to produce the indicated estimates. A standard error of infinity is necessary to produce the least-squares point, and a standard error of zero produces the constrained least-squares estimates. From our perspective, a SIGMA1 in the neighborhood of .1 seems sensible. The long run elasticities, $\Sigma_1^5 \beta_i$ and $\Sigma_6^{10} \beta_i$, are likely to have absolute values in the range of .5 to 2, and it is highly doubtful that neighboring coefficients would differ by more than .2.

A graphical description of the evidence about $\Sigma_6^{10} \beta_i$ is given in Figure 9, and about β_1 in Figure 10. In Figure 6 it is seen that the upper bound is attainable at very low confidence levels, and the lower bound is attainable by a confidence level of .90. The data information is therefore not sufficiently strong that the ambiguity in the prior can be ignored. The contract curve is relatively flat until a probability value of .75 or, equivalently, a SIGMA1 of .28.

Table 4

Contract Curve
 $\sum_1^4 (\beta_i - \beta_{i+1})^2 + \sum_6^{10} (\beta_i - \beta_{i+1})^2$ is small

PROB*	SIGMA1**	β_1	β_2	β_3	β_4	β_5	$\sum_1^5 \beta_i$	β_6	β_7	β_8	β_9	β_{10}	$\sum_6^{10} \beta_i$
0	∞	.20	1.9	-.91	.54	-.34	1.38	-.45	-.47	.45	-.62	.41	-.68
.25	.70	.61	.93	-.04	-.05	-.08	1.38	-.68	-.12	.05	-.14	.12	-.77
.50	.47	.63	.75	.09	-.03	-.07	1.37	-.56	-.15	-.01	-.11	.05	-.79
.75	.28	.57	.57	.19	.04	-.02	1.35	-.39	-.18	-.09	-.14	-.07	-.86
.90	.15	.43	.41	.24	.14	.10	1.33	-.26	-.19	-.17	-.19	-.17	-.98
.95	.08	.33	.31	.26	.22	.20	1.31	-.23	-.21	-.20	-.21	-.21	-1.06
.9696	0	.26	.26	.26	.26	.26	1.30	-.22	-.22	-.22	-.22	-.22	-1.11

*PROB is the classical probability value attaching to the selected ellipsoid.

**SIGMA1 is the prior standard deviation of $\beta_i - \beta_{i+1}$ necessary to produce the indicated point.

Because the posterior standard error is fairly constant, the t-value increases in absolute value as the prior becomes tighter. Regardless of the prior, the estimate is more than one standard error from the origin. In this case the contract curve reveals that a little bit of prior information has a sharp effect on the estimate, but as the prior is further tightened the estimate changes very little. This region of insensitivity corresponds to confidence levels from .15 to .75 and values of SIGMAL from 1.0 to .28. If SIGMAL is further reduced, the estimate of β_1 declines, but the confidence level begins to be unacceptably high. The posterior t value behaves somewhat similarly. A little bit of prior increases the t-value from .45 to 2.0. If the prior is assumed to be very precise, the t-value can be increased to 5.

From this discussion, and from a similar study of the behavior of the other estimates we conclude that a value of SIGMAL equal to .28 produces estimates which are a reasonable compromise between the least-squares estimates and the constrained least-squares estimate. The prior standard error of .28 for $\beta_i - \beta_{i+1}$ represents a rather weak prior. As described above the inferences will change little if the prior is further weakened, but may change considerably if the prior is tightened. But if the prior is tightened the data confidence level grows too high. For SIGMAL = .28 the estimated equation is

$$\begin{aligned}
 M_t = & .60Y_t + .57Y_{t-1} + .19Y_{t-2} + .04Y_{t-3} - .02Y_{t-4} - .39P_t \\
 & (.22) \quad (.17) \quad (.16) \quad (.16) \quad (.20) \quad (.26) \\
 & -.18P_{t-1} - .10P_{t-2} - .14P_{t-3} - .07P_{t-4} - 3.64 \\
 & (.20) \quad (.20) \quad (.22) \quad (.28)
 \end{aligned}$$

If we were forced to select a single equation, this is the one we would choose.

Conclusion

The three examples we have offered illustrate what we believe is a highly useful way of exploring and reporting the sensitivity of inferences to changes in the model. The procedure begins with the choice of a very general model. The models we have in mind include many more explanatory variables than are commonly used in any single estimated equation. The model should be general enough to include as special cases all models which the researcher might like to see estimated with the given data. Adequate degrees of freedom are a concern only for estimating the residual variance σ^2 , an issue which need not concern us here.

Next a family of constraints on the general model is selected. There are likely to be many doubtful variables which have coefficients which are candidates to be set to zero. Other constraints will depend on the setting. The description of the family of models is then completed by selecting a prior variance-covariance matrix for the set of constraints. Constraints which are thought to be quite likely to be approximately true will be assigned relatively small prior variances. Non-zero covariances can be selected when constraints are conceptually interrelated.

The last step in preparation for the data analysis is the selection of a set of issues which are the focus of the inferential exercise. An issue may be a particular coefficient or a linear combination.

The data analysis commences with the computation of the unconstrained least-squares estimates and standard errors of each of the issues and also the computation of the bounds for these issues over all constrained estimates. When these bounds are narrow relative to the sampling standard errors or to anticipated decisions, the process terminates, and it is reported that

uncertainty in the model is essentially irrelevant for all the issues. This is the case for the first example in which the choice of regional dummy variables was shown to have little effect on the estimated coefficients of education, age and age-squared.

When these bounds are wide, it is necessary to analyze and to report the SEARCH diagram which describes (1) bounds constrained to selected confidence ellipsoids and (2) the contract curve. When the constrained bounds are narrow for all reasonable confidence values, again the process may terminate. In this case it is not reported that the uncertainty in the model is essentially irrelevant; rather, it is said that only relatively strong priors can produce estimates which are importantly different from the least-squares estimates. An example is Figure 6, the value of additional floorspace.

Another way to narrow the set of estimates is restrict attention to the contract curve. This curve has meaning only if it is possible to select a quadratic form which measures the prior distance of an estimate from the prior constraint values. For example, doubtfulness of the higher order terms in the functional form example does not seem to imply any special distance function. Shiller's prior for the distributed lag example which, though open to considerable question, nonetheless has a greater intuitive appeal than $\sum_{j=13}^{15} \beta_j^2$ for the functional form example.

Another feature of the distributed lag problem is that the prior scale factor SIGMA1 can be discussed, and the estimates may be finally narrowed to a set of points on the contract curve.

The reporting philosophy which we espouse should now be clear. Because any economic model is open to considerable doubt, the inferences implied by any particular model are of very limited value. The only solution we can see is to report the inferences implied by many different models. A data set determines a mapping from assumptions into inferences: "the mapping is the message." The problem which then arises is how to weigh completeness versus economy in reporting this mapping. We offer in this paper what we believe is both an economical and an informative reporting style.

References

- Chamberlain, G., and E. Leamer (1976), "Matrix weighted averages and posterior bounds," J. Roy. Statist. Soc. B 38, 73-84.
- Dickey, J. M. (1975), "Bayesian alternatives to the F-test and least-squares estimates in the normal linear model," in S.E. Fienberg and A. Zellner (Eds.) Bayesian Studies in Econometrics and Statistics, North Holland, Amsterdam.
- Hoerl, A. E. and R. W. Kennard (1970), "Ridge regression: biased estimation for nonorthogonal problems," Technometrics 12, 69-82.
- Leamer, E. E. (1978), Specification Searches: Ad Hoc Inference with Non-Experimental Data, Wiley, New York.
- Shiller, R. (1973), "A distributed lag estimator derived from smoothness priors," Econometrica 41, 775-788.

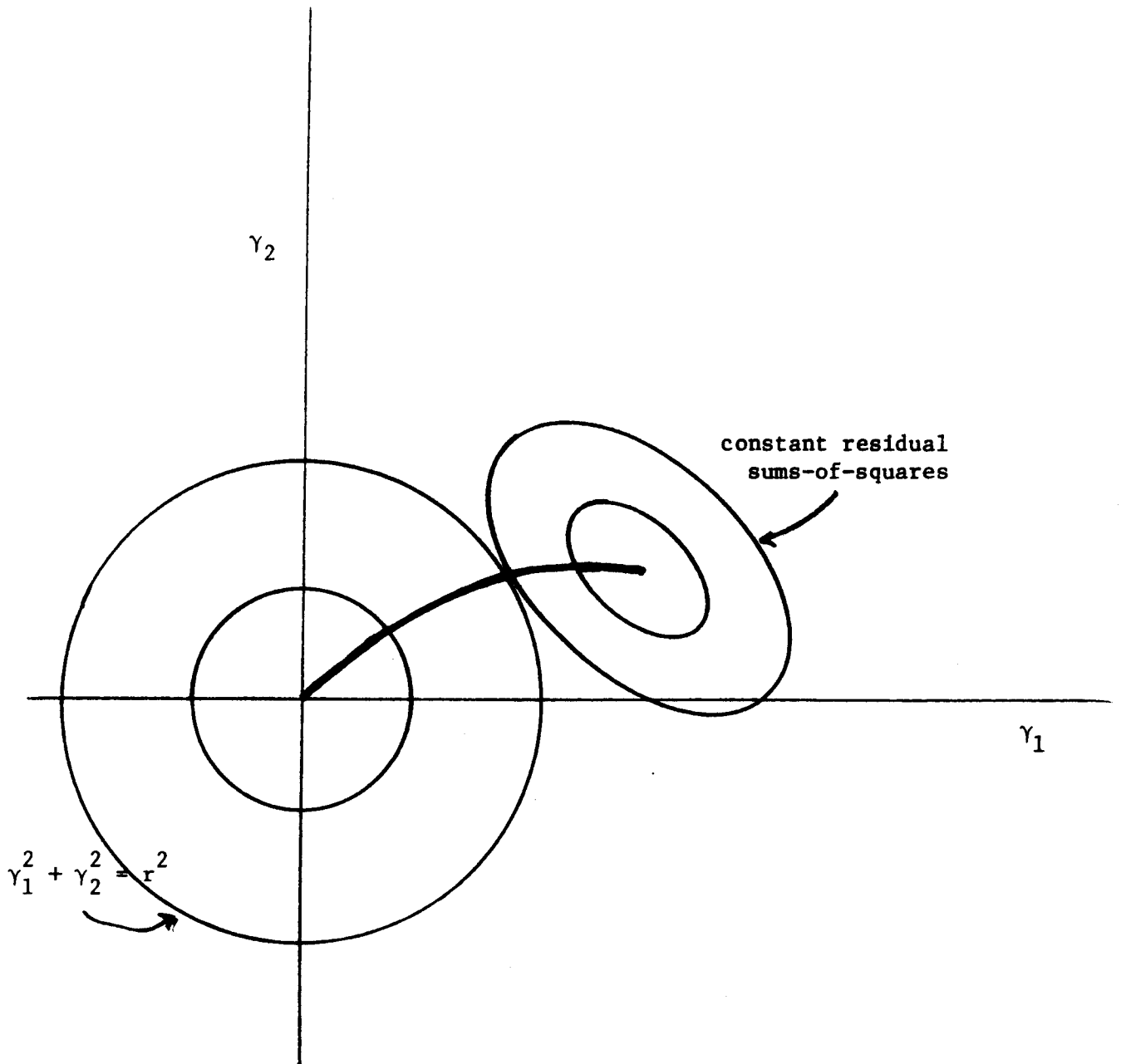


Figure 1 The Information Contract Curve

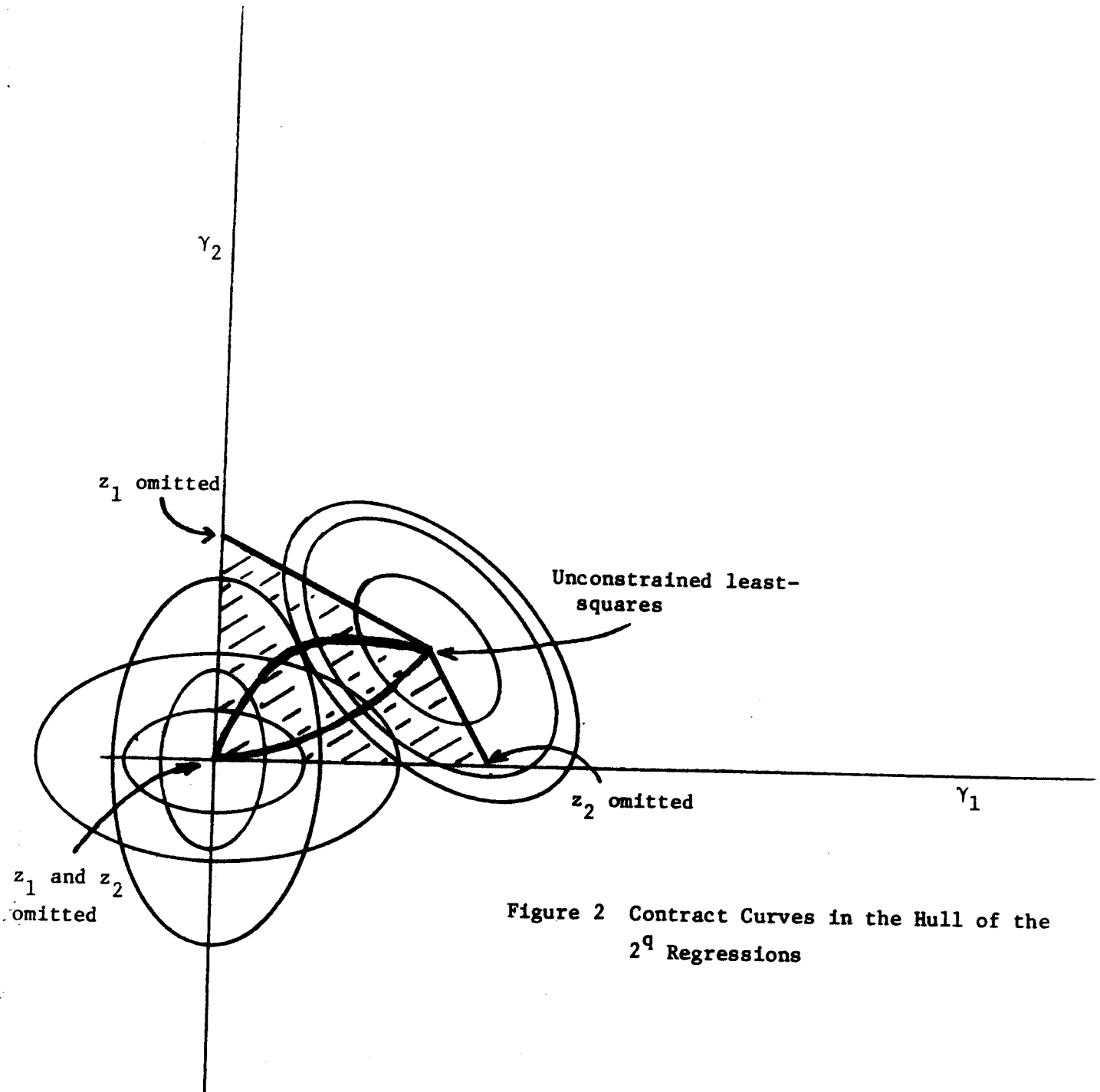


Figure 2 Contract Curves in the Hull of the 2^q Regressions

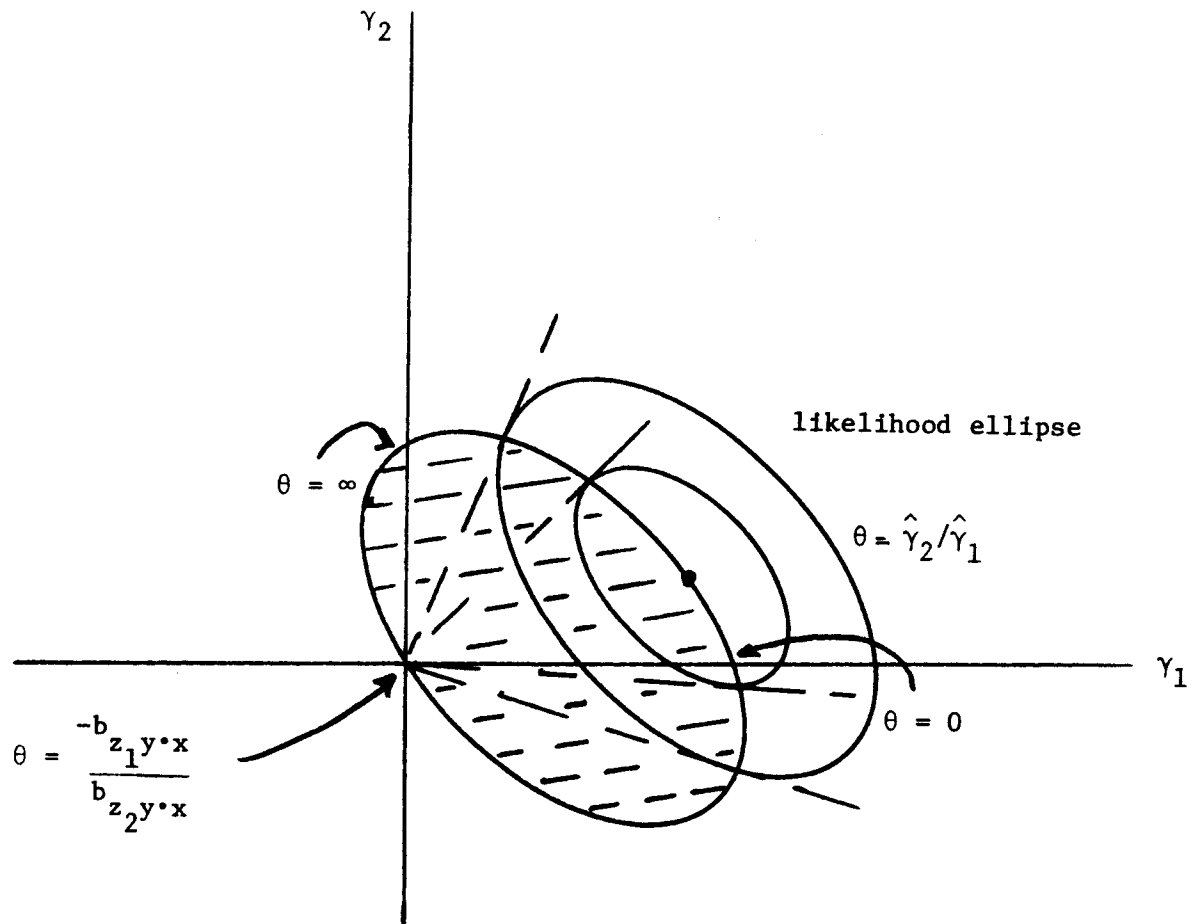


Figure 3 Ellipsoid of Constrained Estimates

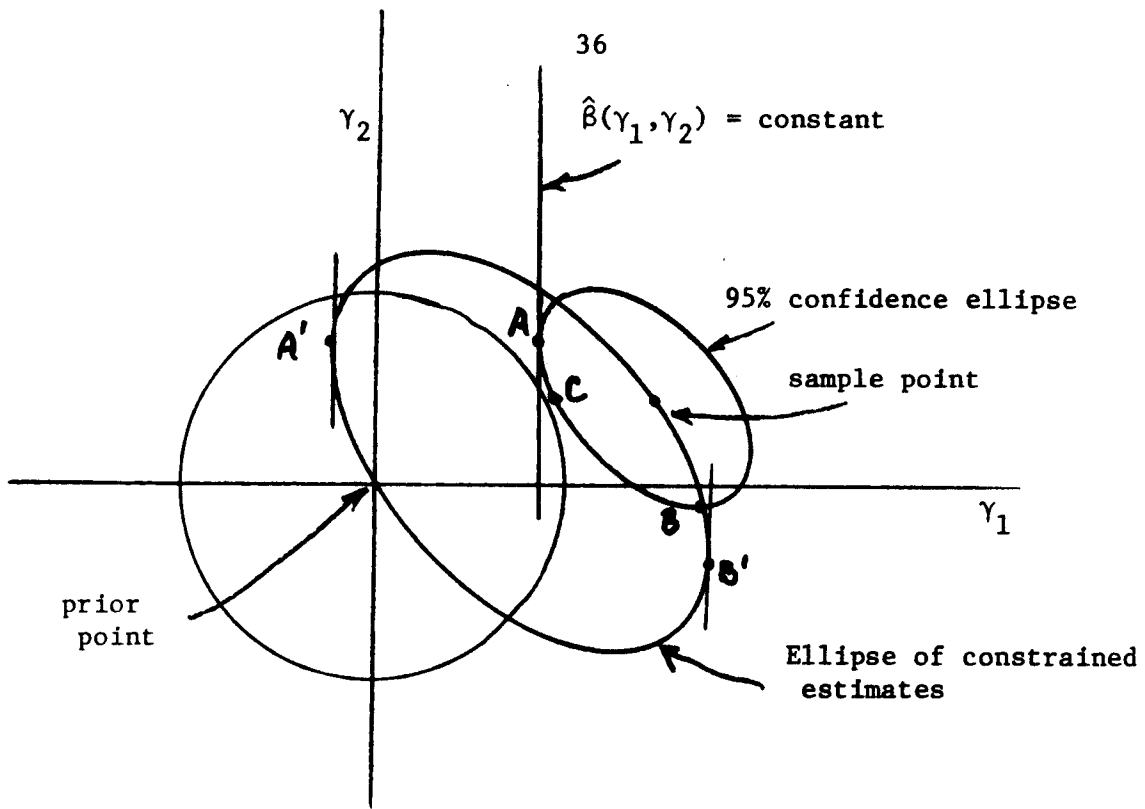


Figure 4 Choice of Points on the 95% Confidence Ellipse

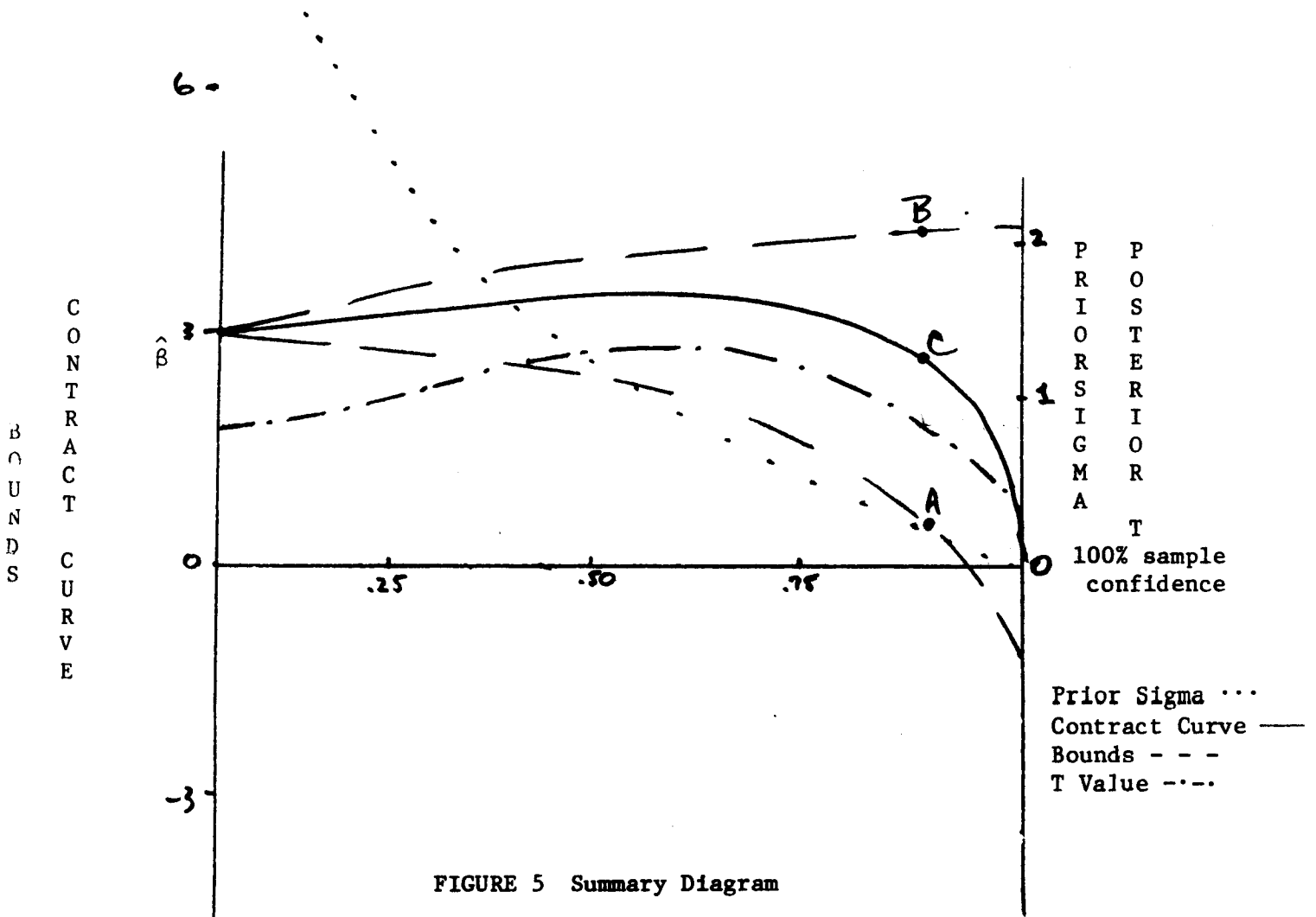


FIGURE 5 Summary Diagram

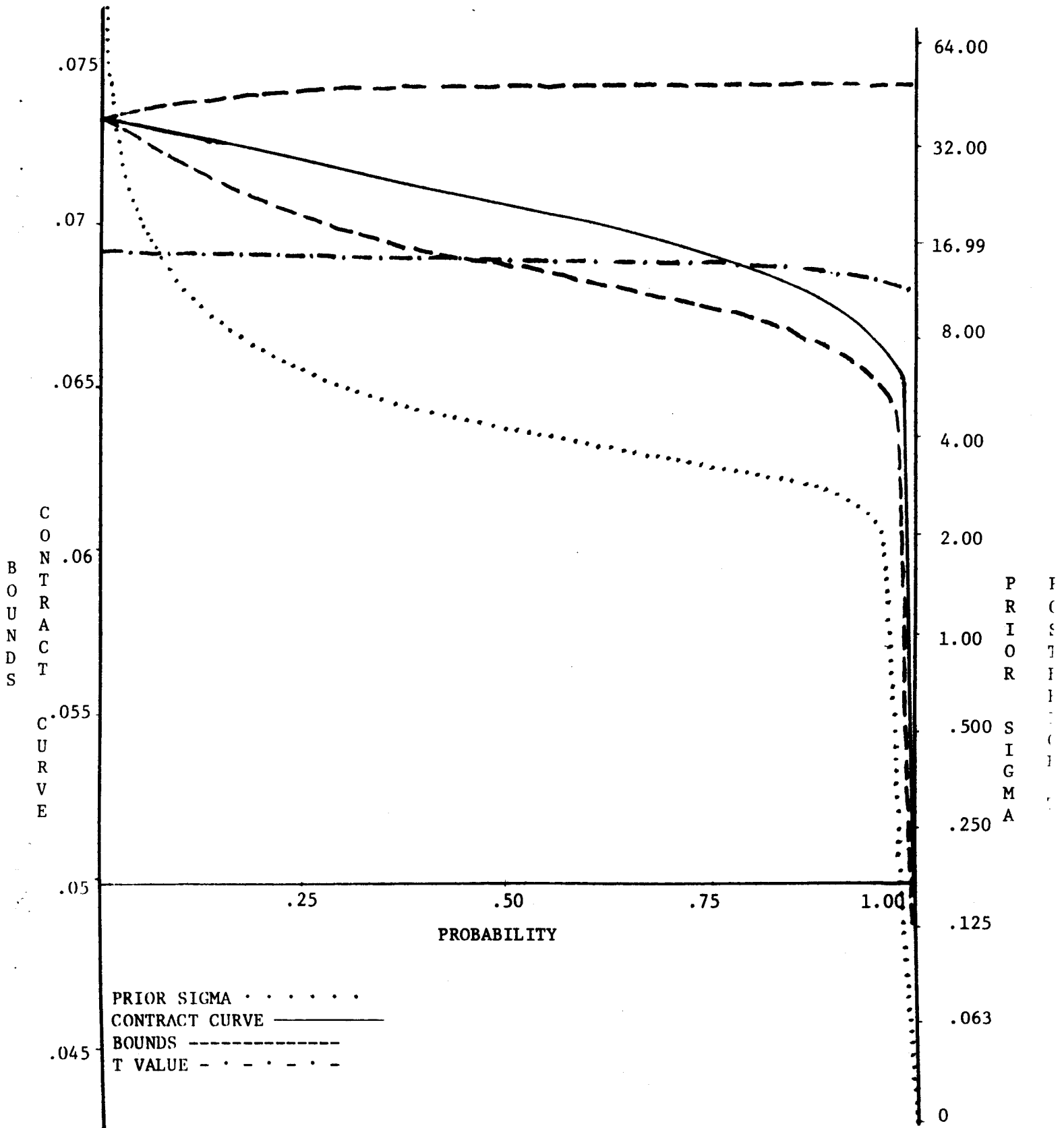


FIGURE 6 Search Graph for Value of 500 sq. ft. of floorspace

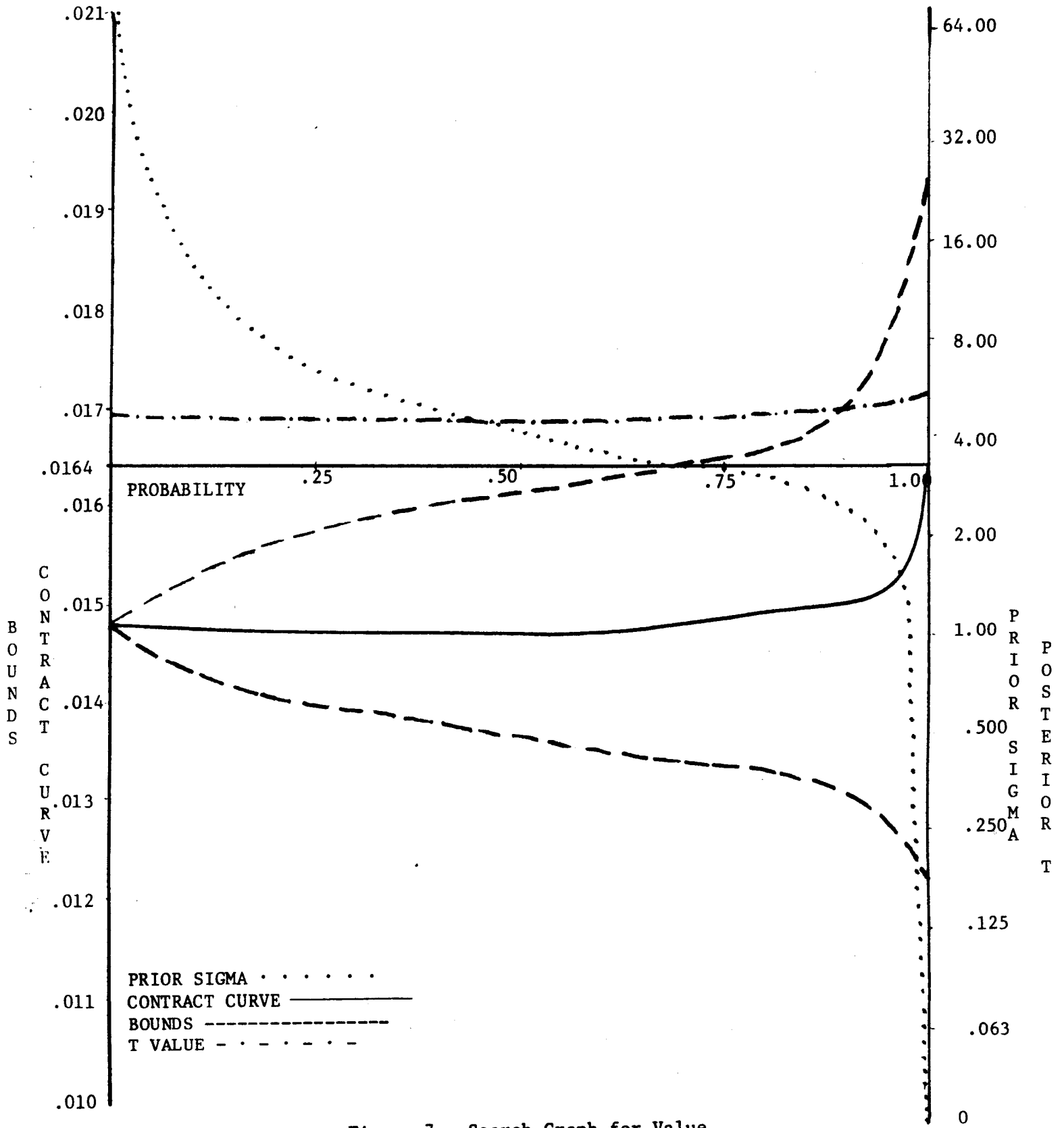


Figure 7 Search Graph for Value of Additional Room

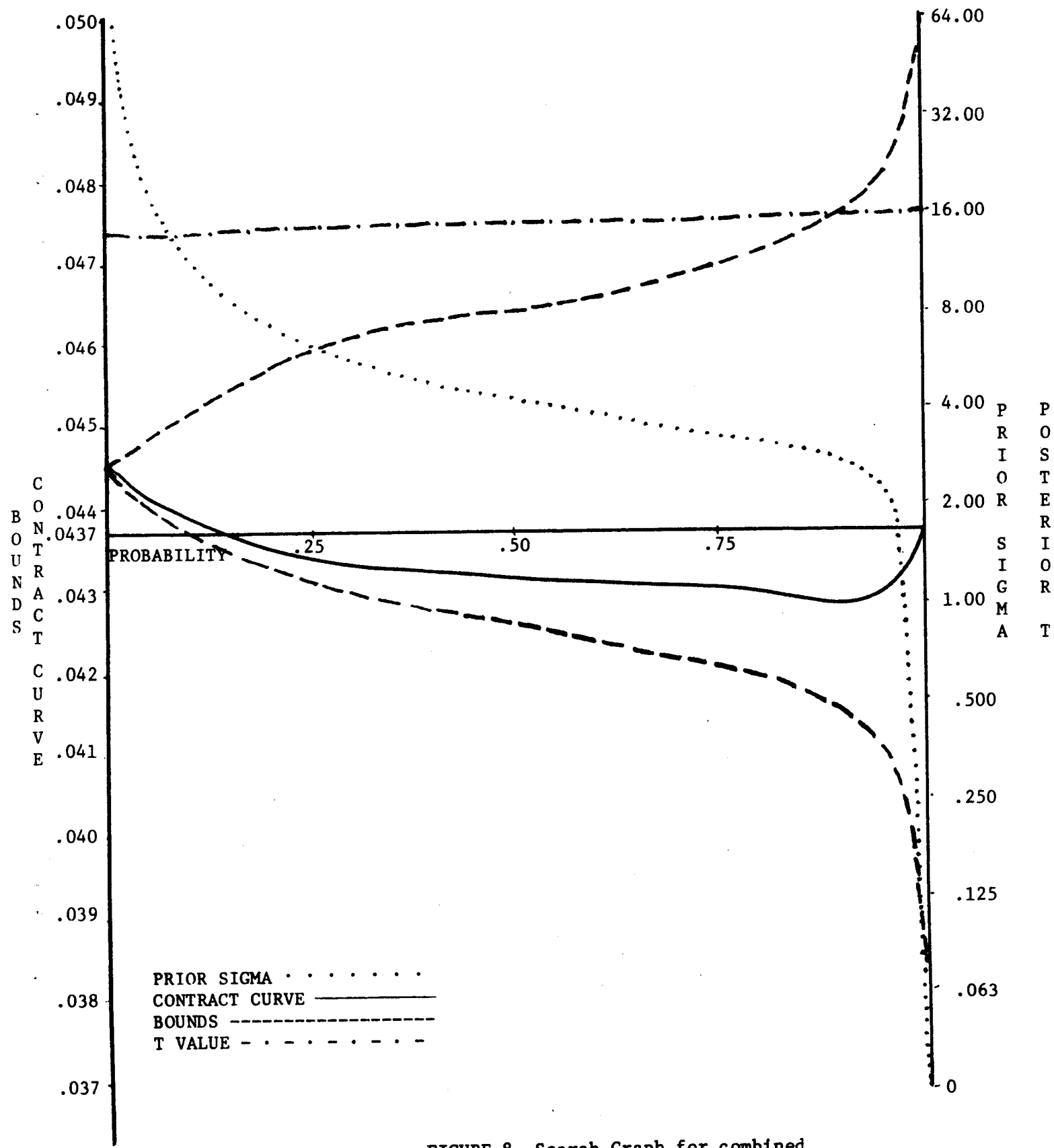


FIGURE 8 Search Graph for combined value of 300 sq. ft. of floor-space and one room

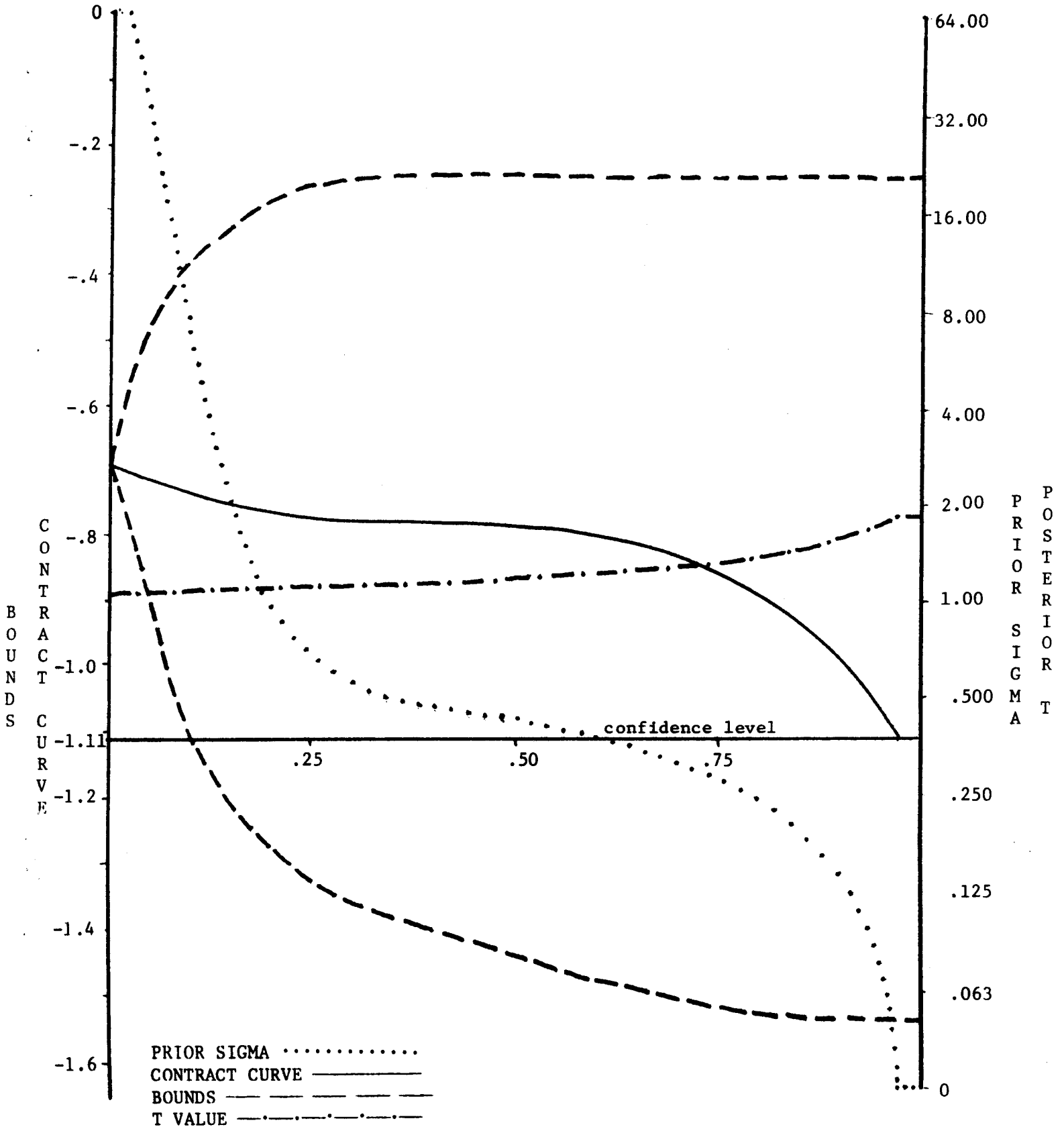


FIGURE 9 SEARCH graph for $\sum_6^{10} \beta_1$, distributed lag problem.

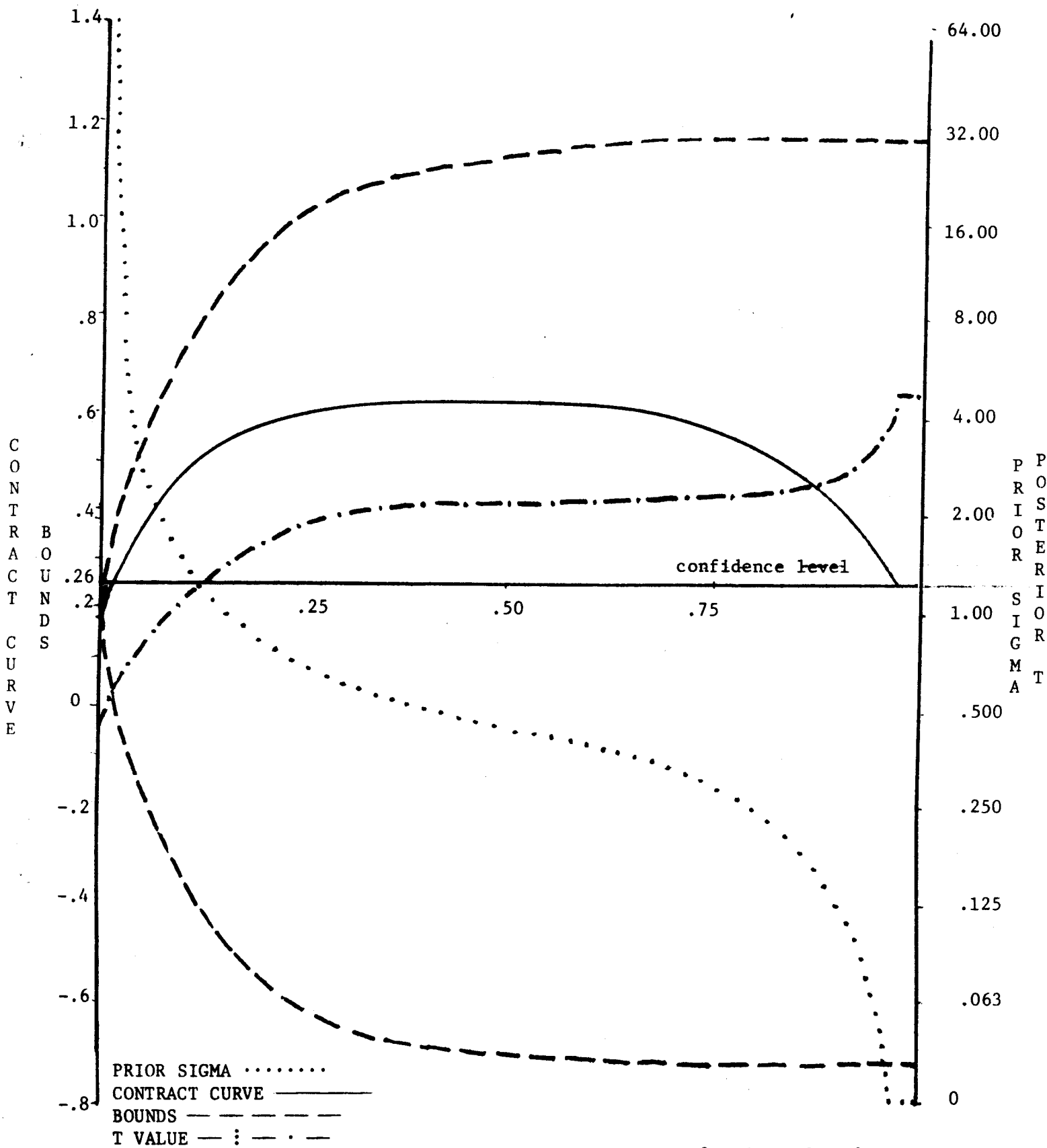


FIGURE 10 SEARCH graph for β_1 , distributed lag problem.