Thesis Submitted for the Degree of

Doctor of Philosophy

1992

ROBUSTNESS OF ESTIMATION BASED ON EMPIRICAL TRANSFORMS.

Edward Campbell

ABSTRACT.

The robustness of certain model-fitting procedures, based on statistical transforms, is investigated using the Influence Function. Our discussion is in two parts. In the first, we focus on estimating the parameters of particular distributions, given independent and identically distributed realizations. We then move on, in the second part, to discuss the fitting of stochastic models.

In this latter context the approach based on transforms, such as the Laplace transform, offers the possibility of explicit parameter estimation. This is in obvious contrast to the more usual situation where only a numerical solution is possible. It was shown by Kemp & Kemp (1987), in a two-parameter example, that only a one-dimensional search was required to produce well-defined estimators. This phenomenon was noted earlier by Morgan (1982), and provides further motivation for transform methods. We generalize the result, and provide an example where a three-dimensional search can be reduced to a line search. With this in mind, we consider the fitting of stochastic models in some detail, employing the standard technique of ordinary least-squares as a bench-mark in this work.

The central theme of this thesis is, however, the robustness of such methods. To this end, we develop a powerful and flexible influence theory in the context of non-indexed random variables. These developments allow us to make concrete statements about the robustness of procedures based on transforms. We show that an analogous treatment is possible for the indexed case, allowing useful qualitative information about parameter estimators to be gathered.

Acknowledgements.

I am grateful to my supervisor, Professor B.J.T. Morgan, for his advice and many helpful comments during the preparation of this thesis. His guidance during this project has been invaluable.

In addition, I am indebted to the obdurate support I have received from Deborah Burge.

This work was funded through a SERC research studentship.

ROBUSTNESS OF ESTIMATION BASED ON EMPIRICAL TRANSFORMS.

CONTENTS:

5. ESTIMATION FOR INDEXED RANDOM VARIABLES.

CHAPTER 1:

INTRODUCTION AND REVIEW.

This thesis progresses naturally from the classical context of independent and identically distributed random variables, to consideration of indexed random variables, which are typically time-ordered. As such, this chapter splits into two parts, beginning with the non-indexed case.

1. Non-Indexed Random Variables.

Procedures based on transforms, such as characteristic and moment generating functions, have been studied for a considerable time. An early motivation seems to have been a desire to fit Stable laws in the modelling of logarithmic changes in stock prices. See, for example, Koutrouvelis (1980) and the original work of Mandelbrot (1963). Since the Stable laws can only, in general, be easily specified in terms of their characteristic functions, then techniques for fitting them to data are based on an estimator of the characteristic function. This estimator is known as the empirical characteristic function, or e.c.f.. In general, the transform estimator is referred to as an empirical transform.

Much of the work to date on transform methods has been concerned with utilizing empirical transforms in estimation and testing, especially in applications where density functions are difficult to obtain. A number of estimation techniques will be studied later. For an introduction to methods of testing statistical hypotheses see Csörgö & Heathcote (1984). A more recent example was provided by Epps & Pulley (1986), who derived a test for exponentiality based on the e.c.f.. The analysis of data observed as sums of independent random variables provides a fruitful area for the application of transform methods. The convolution integral involved in obtaining the density function is rarely straightforward, but the calculation of statistical transforms typically is. An example of such data is provided by Davis & Kutner (1976).

A further example where a probability density is not readily available was provided by Wise (1989), who was concerned with the statistical analysis of Plutonium levels in the environment. Wise gives a model for the measurement of Plutonium in soil, from which an expression is

derived for the Laplace transform of the density function of the overall Plutonium level. It is then natural to employ transform techniques to fit the parameters of the measurement model. This was done by minimizing a weighted least-squares measure between the empirical and theoretical transforms, selecting the transform variables from variance considerations. However, he did experience difficulties here because the transform variables tended to a common value. This phenomenon was noted in the abstract of this thesis, and we will exploit it later. Wise, however, considered this to be less than advantageous and, quite ingeniously and perhaps unwisely, forced a choice away from this diagonal solution.

Despite all this work, very little attention has been given to the robustness of transform methods. This is a pity given that the robustness of certain transform techniques has been noted by a number of researchers, without fully exploring the underlying reasons. Paulson & Nicklin (1983) applied characteristic functions for estimation in linear models, via a technique known as Integrated-Squared-Error (ISE). Heathcote (1977) conducted a general investigation of this method, suggesting that its desirable properties are explained by its interpretation as a density estimation problem. This view of ISE reveals the smoothing implicit in the technique, and was also reported by Bryant & Paulson (1983). The method of ISE has also been applied to the analysis of mixtures with promising results, as investigated by Quandt & Ramsey (1978) and followed-up by Schmidt (1982). This area is reviewed by Titterington et al. (1985) pp 126-33.

A natural means for investigating robustness is the Influence Function, a powerful and flexible tool for gathering qualitative information about statistical quantities. In an ideal situation, the data we observe arise from a particular specified distribution function, F say. In practice, they may actually constitute a sample from some other distribution function F', which is "near" F. A robust estimator is one which performs well at F', whilst not being too inefficient at F. The Influence Function provides a means of investigating this requirement for a proposed estimator. In practice, we cannot know exactly the distribution function from which sampled data arise. We can, however, investigate the performance of an estimator when the data are adjusted slightly. A resistant estimator is one which is not unduly affected by this, whilst performing reasonably well

when no such adjustments take place. The definition of resistance makes no mention of distribution functions, but for all practical purposes the definitions of robustness and resistance are identical. This is an equivalence which we make use of in Chapter 2.

The Influence Function has been applied to a diverse range of problems, many of which are described in the standard texts Hampel et al. (1986) and Huber (1981). See also Critchley (1985), Critchley & Vitiello (1990) and Pack & Jolliffe (1992) for applications in multivariate statistics. Given its strong theoretical and practical background, we employ the Influence Function as the principal means of investigating the robustness of the methods to be discussed in the body of this thesis.

Indeed, the subject matter of Chapter 2 is the Influence Function in general terms. The intention here is to give a self-contained development of this investigative tool, which motivates the formal general definition. Beyond this, we extend the basic definition in a more accessible way than that found in the standard texts. In particular, we consider the Influence Function of a differentiable transform of a statistic having known Influence Function. This result is then extended to deal with collections of statistics combined by means of a differentiable function.

As we see in Chapter 2, the Influence Function may itself be considered as a random variable. As such, its statistical properties are of great interest; these are presented along with a number of illustrative examples which give prominence to an important property of influence functions. This is a standard result of influence theory showing how the asymptotic variance of a statistic may be obtained from its Influence Function. We conclude Chapter 2 by defining an influence-bounded robust statististic as one which possesses a bounded Influence Function.

Given the fundamental material contained in Chapter 2, we are then in a position to investigate the robustness of a number of estimation methods based on empirical transforms. This is the theme of Chapter 3, where we focus on two of the most important of these. Fundamental to this work is the influence behaviour of empirical transforms, which is discussed in general terms. We see that the underlying reason for the robustness of procedures based on them concerns the choice of transform employed. This was noted by Wise (1989) in the context of a method of moment

estimation, which may be viewed as a special case of M-estimation. Wise was then able to appeal to standard theory. In this thesis we approach the robustness of empirical transforms directly to demonstrate that this is a general property of transform methods.

The first technique, that of ISE, has already been mentioned. This is a particularly important member of the general class of integrated-distance methods, so we pay special attention to it. In this context, the transform of choice is usually the characteristic function, so properties of the e.c.f. are of considerable interest. The influence theory developed demonstrates that it is this choice which leads to the observed robustness of estimators derived by the ISE technique. There is an additional attractive feature of ISE, which is the rôle played by an accompanying free parameter, open to choice. We see that we can trade efficiency for extra robustness, and vice-versa, by varying this parameter. In an extended example we are able to obtain a robust estimator of the Normal mean which also achieves very reasonable efficiency in the absence of contamination. In the terms described above, this is an ideal robust estimator.

Perhaps the simplest estimation technique is the method-of-moments, where we simply equate the empirical and theoretical transforms at as many distinct values of the transform variable as there are unknown parameters. Appendix II shows how we may allow these points to become arbitrarily close, eventually taking a common value. It has been noted by Heathcote (1977) that the method-of-moments may be viewed as a special case of ISE. Influence results follow readily via the powerful techniques developed, showing that moment estimators inherit the influence behaviour of the empirical transforms in a linear fashion. An important point to note is that the influence functions of such estimators are easily found, even if no explicit expressions for the estimators are available. Examples show, however, that these estimators can be inefficient in the absence of contamination, sacrificing efficiency for robustness. The method-of-moments lacks the flexibility of ISE to trade-off some of this robustness for greater efficiency, since no free parameter is available.

An important feature of transform methods concerns the selection of the points at which the empirical transform is to be evaluated. Integrated-distance techniques deal with this by

integrating-out the transform variable, whereas the method-of-moments requires that some external criterion be applied. These are often based on variance considerations, and we consider a number of possibilities. There are alternatives however, such as cross-validation and a method known as constrained optimization, which we shall encounter later in this thesis.

The influence techniques developed for investigating the robustness of estimation based on empirical transforms may be applied to a diverse range of problems. This is demonstrated at the end of Chapter 3, where we show that influence functions for estimators defined by very general estimating equations can be found, and that extensions are possible to non-standard transforms and a class of location-scale families of distributions.

In Chapter 4 we investigate some of the practical consequences of the theory developed in the previous chapter, in the context of the Lagged-Normal distribution. The main reason for this choice is that its density function does not exist in closed form. However, since it may be represented as a convolution, an expression for its moment generating function, for example, does. The emphasis here is placed on moment estimation, although we also consider the influence properties of ISE briefly. In employing the method-of-moments we seek to obtain explicit parameter estimators. We conduct an investigation of resulting influence and efficiency properties, in comparison with maximum likelihood. A key step in this process is the calculation of a Jacobian matrix, whose elements are derivatives of the transforms employed with respect to each of the unknown parameters. Such a structure makes this an ideal application of Symbolic Algebra, and we make extensive use of this important tool.

2. Indexed Random Variables.

Having discussed transform methods for independent and identically distributed random variables, it is natural for us to try to relax these assumptions. This brings us into contact with the broad area of stochastic modelling. The typical approach to fitting such models is to minimize, with respect to the unknown parameters, some measure of discrepancy between the observed and fitted observations. Even in relatively simple models, and applying a straightforward discrepancy measure, explicit parameter estimators are unlikely to be available. As such, we have usually to

apply some "black-box" numerical algorithm, which we trust to be reliable. Clearly if some alternative approach could produce explicit, reasonably efficient, estimators then it would be worthy of consideration. Transform procedures would seem to have some potential here and, if nothing more, could provide a useful means of yielding starting values for more complex techniques.

In attempting to fit stochastic models using transforms we bear in mind, in view of the above discussion, the need for simplicity. By implication then, we employ the method-of-moments, as did Leedow & Tweedie (1983) in the modelling of fish growth. They adopted an approach based on Laplace transforms to estimate the parameters of a particular growth curve model, obtaining explicit parameter estimators. For a particular set of data, consisting of growth information on both male and female fish, a revealing comparison was made with three standard techniques for which explicit parameter estimation is not possible. For male fish, two of the methods failed to converge, whilst for female fish two of them produced ridiculous parameter estimates. The reason for this was identified to be an influential observation. Of the four methods studied, only the transform-based technique produced sensible estimates for both male and female fish. This is an interesting observation in the light of the work of this thesis on non-indexed random variables, which suggests that we might expect transform-based estimators to possess good robustness properties. Evaluating the validity of this claim is a key theme of this thesis.

The possibility of obtaining explicit estimators has also been noted by Morgan (1982) and Laurence & Morgan (1987), stemming from the flexibility to choose which aspect of the process to transform. Judicious choice of this quantity can yield explicit estimators attainable. In the case of Laurence & Morgan (1987), this amounted to transforming the survivor function.

If we were to relax the need for simplicity, then ISE might be considered as an estimation technique. This was the approach of Paulson & Nicklin (1983) to fitting linear models. The standard approach is to minimize the sum of squared residuals with respect to the unknown parameters. Paulson & Nicklin replaced the squared residuals by a measure of distance between the theoretical and empirical characteristic functions. These are then summed to give a transform-based measure of the discrepancy between the model and the observations. This technique has

the detraction that strong assumptions must be made as to the nature of the error structure. In order to obtain a theoretical characteristic function, Paulson & Nicklin assumed independent and Normally distributed errors. It is unrealistic to expect these assumptions to be justified in general, and these are not necessary when moment estimation is employed.

Beyond the purely statistical interest in transform methods, there is undoubtedly a need to develop such techniques because transforms occur so often in many scientific disciplines. For example, Alihassan et al. (1992) required the solution of a partial differential equation in the context of Bio-Medical Engineering. This solution was then to be fitted to experimental data by least-squares. However, the differential equation was most easily solved by the method of transforms. The following quote from their paper is instructive:

"Although the estimation procedure could be done in the frequency domain, transformation from the time domain of noisy data with erroneous sample points .... can produce a significant bias .... "

This was their justification for inverting the transform before forming the least-squares objective function. Their particular application employed a Fourier transform, which is essentially a characteristic function. The work of this thesis suggests that the e.c.f. is a robust estimator, in the terms defined earlier. Statistically speaking, then, the above claim by Alihassan et al. may be unjustified, and an empirical Fourier transform could be a suitable quantity to employ given "noisy data". The unknown parameters might then be estimated by ISE, say, avoiding the need to invert a transform.

There are many other applications that could be cited, ranging from the study of queues to the modelling of epidemics. Transforms also arise naturally in Time Series, through the spectral density function. Indeed, Scientists often collect data observed in the frequency domain, which may well be a more familiar setting to them than the time domain. Statisticians should therefore be concerned with developing suitable techniques for analysing such data.

The transform approach to fitting stochastic models is discussed in Chapter 5. Our consideration of this topic is very general at first, employing a Quantal Assay model to illustrate the concepts involved. We argue in this thesis that the Laplace transform is a good choice over a wide variety of applications, for a number of reasons. Firstly, our work on non-indexed random variables suggests that an empirical Laplace transform will possess good robustness properties. Secondly, many stochastic models are often defined by mixtures of exponentials, as is the growth model mentioned earlier. Given this, the Laplace transform will often take a relatively simple form, and so is a natural quantity to consider. A detailed investigation is conducted in this chapter of estimation for a particular member of the class of compartment models, with means described by a system of linear differential equations giving rise to a model of the form noted above.

We focus on the One-Compartment Open model, which Rodda et al. (1975) considered to be "probably the most frequently used model for the pharmacokinetic evaluation of drugs". There is some evidence that this is still the case. Macheras et al. (1992) were concerned with determining the fraction of drug remaining in the body at the time of peak concentration, an important quantity for a medical practitioner to know. Wald et al. (1991), in a study of healthy males, employed a simple compartment model to describe the movement of Basophils (a type of white blood cell) between the blood stream and extravascular sites. The question of interest was how this movement was related to exposure to a variety of steroid. Urfer (1992) used a similarly simple compartmental system to model the uptake of chemical pollutants from the environment by living organisms. These examples demonstrate the wide spectrum of problems to which compartment models are applicable, yielding useful information in many disciplines.

The typical approach to fitting compartment models is via non-linear least-squares. Rodda et al. (1975) were concerned with the robustness of this procedure, finding that it is sensitive to outlying observations. They examined the response of the least-squares estimators to particular outlier patterns, chosen to represent what might be reasonably expected in practice. It was discovered that these outliers could produce effects that would alter clinical decisions. An Ordered Simultaneous Estimation Procedure (OSEP) was developed as a more robust alternative,

and is discussed in greater detail in Chapter 6. Following this work, Frome & Yakatan (1980) re-emphasized the need for resistant estimates in this field. Rather than obtaining estimates by least-squares, they suggest two alternatives. The first is to minimize the sum of absolute deviations between the observed and expected observations. Alternatively, weighted least-squares can be applied with weights chosen to yield a robust procedure. This approach has the disadvantage that this choice is not unique, and the relative merits of different choices should be investigated.

Before we can obtain parameter estimates via the transform approach we must first construct an empirical transform, and then select the values of the transform variable at which it is to be evaluated. In the case of independent and identically distributed random variables, transform construction was trivial. In the present context it is rather less so, and two competing methods emerge. The first of these is based on Gauss-Laguerre quadrature, allowing an element of experimental design to take place. The second method follows from a Riemann-sum approximation to the Laplace transform integral. This latter approach allows us to analyse data not collected from a quadrature-based design. An important function of Chapter 5 is to compare both of these methods.

Since we employ the method-of-moments for parameter estimation, some external criterion is required to select the values of the transform variable at which the empirical and theoretical transforms are to be equated. We use least-squares as a simple means to do this, and compare transform techniques with ordinary least-squares as methods of parameter estimation. This comparison takes the form, in large part, of a simulation study. However, we also investigate the efficiency of transforms relative to ordinary least-squares via more theoretical means.

Having discussed a number of computational and theoretical properties of estimation based on transforms, we move on to consider the robustness of some of these methods in Chapter 6. Once again, we choose the Influence Function as the principal investigative tool. However, it is not immediately clear how its basic definition may be applied directly in this rather novel setting. As such, we follow the development of methods for non-indexed random variables described in Chapter 2 by first introducing a sensible definition of empirical, sample-based, influence.

Following the development of a reasonable approach to measuring influence, we show that it yields the same answer as standard techniques when applied to linear models having fixed effects. The standard approach, as presented by Cook & Weisberg (1982) for example, is a considerably more involved and notably less efficient means of obtaining influence functions than that developed in this thesis.

In order to find influence functions for moment estimators it is necessary, as for the non-indexed case, to develop the elementary theory a little further and this is done in a directly analogous fashion to that of Chapter 3. We are then in a position to discuss the results so obtained for a more general problem than that considered in Chapter 5, showing that useful qualitative information about the transform-based estimators can be easily discerned. As before, it is important to compare the different methods of transform construction. We complete Chapter 6 by examining in detail the influence properties of the transform estimators of Chapter 5 versus those of the least-squares estimators.

CHAPTER 2:

INFLUENCE METHODS FOR NON-INDEXED RANDOM VARIABLES.

1. Introduction.

In Chapter 1 the Influence Function was proposed as a natural tool for investigating the robustness properties of a given statistic. The initial objective of this chapter is to demonstrate how the Influence Function can be adapted to the transform setting. To do this, we develop some general rules for calculating Influence Functions. We are then in a position to understand why, and under what conditions, estimators based on transforms should exhibit robust behaviour.

In section 2 we justify the formal definition of the Influence Function (IF) via the device of empirical influence. Once it is defined, we extend the IF to more complicated and realistic situations. Section 3 is primarily concerned with describing some useful parameters of the IF. In this work we are primarily concerned with determining whether or not the IF is bounded. As such, we do not go into too much detail on these; the interested reader is directed to Barnett & Lewis (1983) pp 140-41.

2. Derivation of the Formal Influence Function.

We have defined the concept of robustness. In practical terms, the robustness of a statistical procedure can only be judged by exploiting the practical equivalence between robustness and resistance. As such, we seek to measure the effect of varying the sample values on a given statistic.

We can express this formally as follows: given a random sample $X_1, \ldots, X_n$ from a distribution F, empirical influence is the effect on $\hat{T}(X_1, \ldots, X_n)$ of adding an extra observation, X say. The following example illustrates the concept.

Example 2.1.

Let $\hat{T}(X_1, \ldots, X_n) = T(\hat{F}) = \bar{X}$; where $\hat{F}$ denotes the sample cdf. Thus,

$$T(\hat{F}) = \int x \, d\hat{F}(x) .$$

Then, letting $\tilde{T}$ denote the "disturbed" statistic

$$\tilde{T}(\underline{X},X) = (n+1)^{-1}(n\bar{X}+X) ,$$

in obvious notation. The effect on $\hat{T}$ can be measured by the difference

$$\tilde{T}-\hat{T} = (n+1)^{-1}(X-\bar{X}) .$$

Not surprisingly, this is proportional to $(n+1)^{-1}$, the proportion of contamination in the sample. Standardizing for this, we define the Empirical Influence of X on $\hat{T}$ to be

$$EI(X;\hat{T}) = (n+1)\{\tilde{T}-\hat{T}\} \tag{2.1}$$

Example 2.1 - continued.

The above definition yields

$$EI(X;\bar{X}) = X-\bar{X} .$$

It is clear from this that $\bar{X}$ can be made to exceed any finite bound for X large enough.

Example 2.2.

Let $T(\hat{F})$ be the sample variance S^2 for a distribution with unknown mean and variance. Then, in obvious notation

$$nS^2 = (n-1)S^2 + n\bar{X}^2 + X^2 - (n+1)^{-1}(n\bar{X}+X)^2 .$$

Thus,

$$\tilde{S}^2 - S^2 = (n+1)^{-1}(X-\bar{X})^2 - n^{-1}S^2 ,$$

so that

$$EI(X;S^2) = (X-\bar{X})^2 - (1+1/n)S^2 .$$

The empirical influence is again unbounded, but is a quadratic function of the added observation.

The empirical influence depends on X, the sample size n, the statistic $\hat{T}$ and, in general, on the underlying distribution F. The asymptotic equivalent is especially useful since it removes the explicit dependence on a particular sample. Thus, we define an Influence Function to be

$$IF(X;\hat{T}) = \lim_{n\to\infty} EI(X;\hat{T}) . \tag{2.2}$$

Example 2.1 - continued.

The influence function for $\bar{X}$ is

$$IF(X;\bar{X}) = \lim_{n\to\infty}(X-\bar{X})$$
$$= X-\mu,$$

replacing $\bar{X}$ in the limit by the population parameter.

Example 2.2 - continued.

$$IF(X;S^2) = \lim_{n\to\infty}\{(X-\bar{X})^2-(1+1/n)S^2\}$$
$$= (X-\mu)^2-\sigma^2 .$$

Now, we have defined the influence function to be

$$IF(X;\hat{T}) = \lim_{n\to\infty}(n+1)\{\tilde{T}-\hat{T}\} \tag{2.3}$$
$$= \lim_{n\to\infty}(n+1)[T\{(1-(n+1)^{-1})\hat{F}+(n+1)^{-1}A\}-T\{\hat{F}\}] , \tag{2.4}$$

where A denotes the atomic distribution allocating all probability at X. Writing $\varepsilon=(n+1)^{-1}$, we find

$$IF(X;\hat{T}) = \lim_{\varepsilon\to 0}\varepsilon^{-1}[T\{(1-\varepsilon)\hat{F}+\varepsilon A\}-T\{\hat{F}\}].$$

For n large enough, $\hat{F}$ may be replaced by its asymptotic dual. Doing this, we arrive at the formal definition of an influence function

$$IF(X;\hat{T}) = \lim_{\varepsilon\to 0}\varepsilon^{-1}[T\{(1-\varepsilon)F+\varepsilon A\}-T\{F\}] . \tag{2.5}$$

We may write this as

$$IF(X;\hat{T}) = \frac{\partial}{\partial\varepsilon}[T\{(1-\varepsilon)F+\varepsilon A\}], \tag{2.6}$$

evaluated at $\varepsilon = 0$. Thus it is clear that $IF(X;\hat{T})$ measures the instantaneous perturbation of $\hat{T}$ due to an observation X.

To conclude this derivation, note that particular emphasis has been placed on the approach through empirical influence for two major reasons. Firstly, we avoid presenting a purely theoretical, unmotivated definition. The examples serve to retain the link with practical considerations. Secondly, and most importantly, we find later that this approach is helpful when indexed random

variables are under consideration. A rigorous discussion of influence functions and their properties is provided by Huber (1981).

Two results follow from (2.6) immediately :-

Corollary 2.1.

Let $\hat{Y} = H(\hat{T})$, where H denotes some differentiable function. Then

$$IF(X;\hat{Y}) = \frac{\partial H}{\partial T} IF(X;\hat{T}). \tag{2.7}$$

Proof

By definition,

$$\begin{aligned} IF(X;\hat{Y}) &= \frac{\partial}{\partial \varepsilon} H[T\{(1-\varepsilon)F + \varepsilon A\}] \\ &= \frac{\partial H}{\partial T} \frac{\partial}{\partial \varepsilon} [T\{(1-\varepsilon)F + \varepsilon A\}] , \end{aligned}$$

where all derivatives are evaluated at $\varepsilon=0$.

□

Example 2.3.

Suppose that we wish to find $IF(X;\hat{Y})$, where $\hat{Y} = \log(\bar{X})$. Then $H(\hat{T}) = \log(\hat{T})$, so that

$$\begin{aligned} IF(X;\hat{Y}) &= T^{-1} IF(X;\hat{T}) \\ &= \mu^{-1}(X-\mu) , \end{aligned}$$

replacing $\bar{X}$ by μ in the limit.

Corollary 2.2.

Let $\underline{T} = (T_1, \ldots, T_p)'$, then in obvious notation

$$IF(X;\hat{Y}) = \sum_{i=1}^{p} \frac{\partial H}{\partial T_i} IF(X;\hat{T}_i) . \tag{2.8}$$

Proof

By definition,

$$IF(X;\hat{Y}) = \frac{\partial}{\partial \varepsilon} H[\underline{T}\{(1-\varepsilon)F + \varepsilon A\}]$$
$$= \sum_{i=1}^{p} \frac{\partial H}{\partial T_i} \frac{\partial}{\partial \varepsilon} [T_i\{(1-\varepsilon)F + \varepsilon A\}]$$

□

This provides an attractive extension of the previous result.

Example 2.4.

We use the technique derived above to find the influence function of $\hat{T} = S^2/\bar{X}^2$. That is, the square of the sample coefficient of variation. Write $\hat{T} = H(\hat{T}_1, \hat{T}_2)$, where $H(T_1, T_2) = T_1 T_2^{-2}$ so that

$$IF(X;\hat{T}) = \frac{\partial H}{\partial T_1} IF(X;\hat{T}_1) + \frac{\partial H}{\partial T_2} IF(X;\hat{T}_2)$$
$$= T_2^{-2} IF(X;\hat{T}_1) - 2T_1 T_2^{-3} IF(X;\hat{T}_2)$$
$$= \mu^{-2}[(X-\mu)^2 - \sigma^2] - 2\sigma^2\mu^{-3}(X-\mu).$$

The same result may be derived via empirical influence, but the result is far more easily obtained in the manner described above. Note that it is obvious from the above that an extension to non-identically distributed random variables is readily available. We do not pursue such an extension here since it is not of importance to this thesis.

3. Properties of Influence Functions.

This section divides naturally into two distinct parts. In the first, we look at some statistical properties of the IF. Secondly, we define some useful parameters of the influence function and discuss what is meant by robustness in the influence context.

3.1. Statistical Properties of the IF.

In this section we discuss the theoretical foundation for the study of influence functions. The core material presented here has been abstracted from several sources. Barndorff-Nielsen &

Cox (1989) pp 189-192 set their discussion in the context of generalized notions of differentiation. Cox & Hinkley (1986) pp 345-46 provides a good introduction to the theory of influence functions. Greater detail is provided by the standard references Huber (1981) and Hampel et al. (1986).

Now, it is readily shown that $\hat{F}(z)$ is uniformly consistent for the true distribution $F(z;\Theta)$. Therefore, provided T(.) is continuous, $\hat{T}=T(\hat{F})$ is consistent for $T(F)$. Note that $\hat{T}$ is by definition Fisher consistent.

The behaviour of $\hat{T}$ relative to Θ depends, then, on the behaviour of $T(\hat{F})-T(F)$, where $\hat{F}(z)$ is a consistent and asymptotically normal estimate of $F(z;\Theta)$. When $F(z;\Theta)$ is discrete, we can make a Taylor expansion of $T(\hat{F})-T(F)$ in terms of derivatives of $T(.)$.

However, where $F(z;\Theta)$ is allowed to be continuous, a more general Taylor expansion is required in terms of *von-Mises derivatives*. Here the fundamental concept is that $T[(1-\varepsilon)F+\varepsilon H]$ is a differentiable function of ε, and the first derivative of $T(.)$ w.r.t. F at z, denoted by $IF(X;\hat{T})$ is defined by

$$\lim_{\varepsilon\to 0}\frac{T[(1-\varepsilon)F+\varepsilon H]}{\varepsilon}=\int IF(z;\hat{T})dH(z). \tag{2.9}$$

Alternatively, $IF(X;\hat{T})$ is given explicitly from (2.9) by choosing H as the atomic distribution allocating all probability at X.

Result 2.1.

By definition (2.9),

$$\int IF(z;\hat{T})dF(z;\Theta)=0. \tag{2.10}$$

We therefore obtain the expansion

$$\begin{aligned} T[(1-\varepsilon)F+\varepsilon H] &= T(F)+\varepsilon\int IF(z;\hat{T})d[H-F](z)+O(\varepsilon^2)\\ &= T(F)+\varepsilon\int IF(z;\hat{T})dH(z)+O(\varepsilon^2), \end{aligned} \tag{2.11}$$

on applying the above result. Thus, with $\varepsilon=1$ and $H(z)=\hat{F}(z)-F(z;\Theta)$ we obtain

$$T(\hat{F})=T(F)+\int IF(z;\hat{T})d\hat{F}(z)+O_p(n^{-1}). \tag{2.12}$$

The above follows from (2.10) and by noting that $\hat{F}(z)-F(z;\Theta)$ is $O_p(n^{-1/2})$. Finally,

$$T(\hat{F}) - T(F) = n^{-1} \sum_{j=1}^{n} IF(X_j;\hat{T}) + O_p(n^{-1}), \tag{2.13}$$

given observed data $X_1, \ldots, X_n$ from F. Thus, the random variables $IF(X_j;\hat{T})$ are i.i.d. with zero mean (Result 2.1) and variance given by

Result 2.2.

$$V(\hat{T};F) = \int [IF(z;\hat{T})]^2 dF(z;\Theta).$$

This result gives us an often convenient means for calculating asymptotic variances. Indeed, it is often far more efficient than the delta method for example. Result 2.1 provides a check that a calculated influence function is valid.

By the Central Limit Theorem, we obtain the asymptotic result

$$T(\hat{F}) \sim N(T(F), n^{-1} V(\hat{T};F)). \tag{2.14}$$

We may extend this to the multi-dimensional case as

$$\underline{T}(\hat{F}) \sim MVN(\underline{T}(F), n^{-1} V(\underline{\hat{T}};F)), \tag{2.15}$$

where

$$V(\underline{\hat{T}};F) = \int IF(X;\underline{\hat{T}}) IF(X;\underline{\hat{T}})' dF(X;\Theta), \tag{2.16}$$

with the expectation carried-out element-by-element. Note that

$$IF(X;\underline{\hat{T}}) = (IF(X;\hat{T}_1), \ldots, IF(X;\hat{T}_p))'.$$

Example 2.1 - continued.

Recall that $IF(X;\bar{X}) = X - \mu$, so that

$$n\ var(\bar{X}) \approx \int_{-\infty}^{\infty} (X-\mu)^2 dF(X)$$

$$= \sigma^2.$$

Thus, we see that Result 2.2 can yield exact results as well as approximations. In particular, the expression so obtained will be exact for statistics which may be written directly as functionals of F.

Example 2.5.

We discuss here an extended example to investigate the quality of the approximation we can expect to achieve using Result 2.2. It differs from the other examples of this chapter in that observations follow a discrete distribution, the Negative-Binomial.

We investigate a particular estimator proposed by Kemp and Kemp (1987), chosen to be explicit. The Negative-Binomial is parameterized so that the P.G.F. is

$$\Pi(z) = [1+q(1-z)]^{-k},$$

for parameters $k,q > 0$. Thus, this distribution has mean $\mu = kq$ and variance $\sigma^2 = kq(1+q)$. We assume k is known, and consider the Kemp & Kemp estimator of q for low (less than two) mean; we denote this by q^*. Explicitly, given a random sample $X_1, \ldots, X_n$ we calculate

$$q^* = (a\bar{X} - \bar{X}_w)/((1-a)\bar{X}_w)$$

where $a = (k+1)/(k+2)$ and

$$\bar{X}_w = \sum X_j a^{X_j} f_{X_j} \; / \; \sum a^{X_j} f_{X_j} \; ;$$

f_X representing a relative frequency.

Thus we may write $q^* = H(T_1, T_2, T_3)$, where

$$H(T_1, T_2, T_3) = a(1-a)^{-1} T_1/(T_2/T_3) - (1-a)^{-1} \; ,$$

in obvious notation. Applying the techniques of this chapter, we calculate

$$IF(X;q^*) = a(1-a)^{-1} \sum_{i=1}^{3} \frac{\partial H}{\partial T_i} IF(X;\hat{T}_i)$$

$$= a(1-a)^{-1} \{T_3/T_2 IF(X;\hat{T}_1) - T_1 T_3/T_2^2 IF(X;\hat{T}_2) + T_1/T_2 IF(X;\hat{T}_3)\}.$$

Now,

$$T_1(F) = \mu \; ,$$

$$T_2(F) = \int X a^X dF(X)$$

$$= a \frac{\partial \Pi(z=a)}{\partial z}$$

$$\Rightarrow T_2(F) = a\mu[1+q(1-a)]^{-(k+1)} \; ,$$

$$T_3(F) = \int a^X dF(X) = [1+q(1-a)]^{-k}.$$

Since these functionals are all of the form

$$T_i(F) = \int t_i(X) dF(X) \quad i = 1,2,3,$$

(2.6) yields

$$IF(X;\hat{T}_i) = t_i(X) - T_i(F).$$

Bringing these results together, we obtain an expression for $IF(X;q^*)$. We concentrate now on the particular example k=5, q=0.2. Using the result

$$n\ var(q^*) \approx \int [IF(X;q^*)]^2 dF(X)$$

we find

$$var(q^*) \approx 3.4/n\ .$$

We can examine this approximation by means of a simulation study. This is especially easy to do since, if X has a Negative-Binomial distribution, then

$$X = \sum_{i=1}^{k} Y_i\ ,\ \ Y_i \sim Geo((1+q)^{-1})$$

in this parameterization. The simulation results are summarized below, obtained from 1000 repetitions in each case.

Table 2.1. Simulation Study of $var(q^*)$.

n	n.var
10	3.03
25	3.08
50	3.11
100	3.16
1000	3.32

We see that the approximation is very good, even for a sample as small as ten.

3.2. Useful Parameters of the IF.

3.2.1. The Gross-Error Sensitivity.

This is typically denoted by γ and is defined to be

$$\gamma(\hat{T}) = \sup_X |IF(X;\hat{T})|.$$

As such, γ measures the largest, in absolute terms, effect a small amount of contamination of fixed size can have on $\hat{T}$. It may therefore be regarded as an upper-bound on the (standardized) asymptotic bias of the estimator.

Examples 2.1, 2.2 - continued.

For both $\bar{X}$ and S^2, $\gamma = \infty$. Thus the effect a contaminant can have on these statistics is unbounded.

3.2.2. The Rejection Point.

Suppose that μ denotes a central location point of the underlying distribution F. Then the rejection point, ρ, is defined by

$$IF(X;\hat{T}) = 0 \quad \text{for} \quad |X-\mu| \geq \rho.$$

That is, observations greater in magnitude than ρ from μ have no effect on $\hat{T}$.

It is possible, and examples of this will be encountered in later work, for $|IF(X;\hat{T})|$ to be small for $|X-\mu|$ large. In this case ρ is infinite, but outliers have little effect on the estimator. Indeed, since outlying observations have similar influence to observations nearer μ, it could be said that outlying observations are brought into the procedure rather than rejected.

3.2.3. The Local-Shift Sensitivity.

We measure here the effect of small fluctuations in the observations, as happens in rounding, grouping and Winsorizing for example. Intuitively, the effect of shifting an observation from X to X' can be measured by $IF(X';\hat{T})-IF(X;\hat{T})$. The standardized effect of this shifting can be described by the slope of the influence function between X and X', motivating the definition

$$\beta(\hat{T}) = \sup_{X \neq X'} |(IF(X';\hat{T})-IF(X;\hat{T}))/(X'-X)|,$$

where $\beta(.)$ denotes the local-shift sensitivity.

3.2.4. The Breakdown Aspect of Performance.

The breakdown point, $b(\hat{T})$, is the smallest proportion of contamination which can cause $\hat{T}$ to exceed all bounds.

An excellent discussion of these quantities is conducted in Hampel et al. (1986). In this thesis we will be primarily concerned with what we term *influence-bounded robustness* (IBR). A statistic possesses this property iff the following condition is satisfied

$$|IF(X;\hat{T})| \leq M, \quad some\ M \in R^{+}. \tag{2.17}$$

This definition of robustness is open to some debate. Indeed, Miller & Halpern (1980) impose the condition that the IF should tend to zero as $|X|$ becomes large. Whilst this is in general desirable, it is perhaps too restrictive in the transform setting. In particular, when characteristic functions are employed in moment estimation, the resulting influence functions are by implication infinitely-oscillating.

Armed with the work of Chapter 2, we are now in a position to investigate techniques for estimation based on transforms. We focus on two of the most important of these in the next chapter.

CHAPTER 3:

ROBUSTNESS OF EMPIRICAL TRANSFORM TECHNIQUES.

1. Introduction.

The IF has been proposed and developed as a tool for investigating the robustness of statistical procedures. We now apply these methods to a study of techniques based on empirical transforms.

It is a feature of transform methods that they really come in three parts. In the first, we construct an empirical transform from observed data to estimate its theoretical partner. Secondly, we must apply some method to extract the information conveyed; two such methods are explored later in this chapter. Lastly, but by no means least, we have to choose the values at which to evaluate the empirical transform. In moment estimation, we must apply some criterion, usually based on variance, to do this. Integrated-distance estimators are formed by combining the second and third steps, thus avoiding the need for some external criterion, an appealing feature of this class of estimator.

In section 2 we formalize our study of empirical transforms and investigate their influence behaviour. Section 3 is devoted to a study of integrated-distance methods, particularly when the measure of distance employed is the square of the l_2 norm. This is known, for obvious reasons, as Integrated-Squared-Error (ISE) where the transform of choice is the empirical characteristic function (e.c.f.). The third section brings perhaps the simplest transform technique, moment estimation, under investigation. We see that this method, although often easy to apply, is lacking in the flexibility offered by ISE. In contrast, the work on moment estimation demonstrates the great range of applications that the influence methods developed in this thesis have. The final section provides further evidence of this. Here we see that it is possible to find influence functions for estimators defined by really quite general estimating equations. The central themes of this chapter are discussed by Campbell (1992).

2. Influence for Empirical Transforms.

Given a distribution function $F(x;\Theta)$, for parameters $\Theta = (\theta_1, \ldots, \theta_p)'$, we may define a transform by

$$G(s;\Theta) = \int g(s,X)dF(X;\Theta). \tag{3.1}$$

Here $g(s,X)$ is known as the *kernel* of the transform (Feuerverger and McDunnough, 1984). For example, choosing $g(s,X) = e^{sX}$ we obtain the moment generating function. For a random sample $X_1, \ldots, X_n$ from F, the empirical transform is defined as

$$\begin{aligned}\hat{G}(s) &= \int g(s,X)d\hat{F}(X) \\ &= n^{-1}\sum_{j=1}^{n} g(s,X_j).\end{aligned} \tag{3.2}$$

Note that, by definition, we have convergence in probability of $\hat{G}$ to G. Aspects of estimating statistical functionals are discussed by Hall & Johnstone (1992).

The influence function of $\hat{G}$ is easily found as

$$\begin{aligned}IF(X;\hat{G}) &= \frac{\partial}{\partial\varepsilon}\int g(s,X)d[(1-\varepsilon)F+\varepsilon A] \\ &= \int g(s,X)d[A-F] \\ &= g(s,X) - G(s;\Theta).\end{aligned} \tag{3.3}$$

This is a general result for functionals which may be expressed in the form (3.1).

It is clear from (3.3) that the influence behaviour, and hence robustness properties, of $\hat{G}$ are determined by the form of the kernel. In the theory of M-estimators, considerable work has been done on the choice of estimating equation required to yield given influence properties. The above result shows that we are able to do this in the transform setting also, the key quantity being the transform kernel.

3. Integrated-Distance Methods.

This method seeks to minimize a measure of distance, $d(\hat{G},G)$ say, between the empirical and theoretical transforms. Before minimizing, the transform variable is integrated-out to leave a quantity involving the data and Θ alone. The ISE estimator, $\hat{\Theta}$ of Θ is found, then, as

$$h(\hat{\Theta}) = \min_{\Theta} \int_{-\infty}^{\infty} d(\hat{G},G)w(s;\lambda)ds. \tag{3.4}$$

It is usual, as above, to include a weight-function in the integral. This is done firstly to ensure integrability and, secondly, to make the integration easier or even possible. Note that there is a free parameter, λ, open to choice. We will investigate the *rôle* it plays in this technique during the course of this chapter.

The most common form for the distance measure is

$$d(\hat{G},G) = |\hat{G}(s) - G(s;\Theta)|^p, \tag{3.5}$$

for $p \in R^+$. The most popular choice seems to be $p=2$, resulting in ISE.

Turning now to the choice of weight-function, the published literature contains references to two major categories. In the first, we choose $w(s;\lambda)$ to be a Heaviside-type function, so that we restrict attention to a finite collection of points. This scheme was followed by Quandt and Ramsey (1978), who employed

$$w(s) = \begin{cases} 1 & \text{if } s \in S \\ 0 & \textit{otherwise.} \end{cases}$$

We concentrate on continuous weight-functions here and return to the subject of discrete weight-functions later in this chapter.

We focus now on a study of ISE and, in this context, the transform of choice is the e.c.f., some properties of which we discuss next.

3.1. The e.c.f..

The e.c.f., denoted by $\hat{\phi}$, has kernel function

$$g(s,X) = e^{isX}, \tag{3.6}$$

so that

$$\begin{aligned} \hat{\phi}(s) &= n^{-1} \sum_{j=1}^{n} e^{isX_j} \\ &= \hat{R}(s) + i\hat{I}(s), \textit{ say.} \end{aligned} \tag{3.7}$$

We see from this that outlying values are trigonometrically reduced. Explicitly, we find by stan-

dard methods that

$$IF(X;\hat{R}) = cos(sX) - \text{Re}\,\phi(s;\Theta)$$
$$IF(X;\hat{I}) = sin(sX) - \text{Im}\,\phi(s;\Theta), \tag{3.8}$$

replacing $\hat{\phi}$ by its theoretical partner ϕ. We see that $\hat{R}$ and $\hat{I}$ are Influence-Bounded-Robust (IBR).

The e.c.f. has found many diverse applications, of which we can mention but a small subset here. In Csörgö & Heathcote (1984) it is employed in tests for independence, symmetry and normality. They also investigate a test for stability, an important application of e.c.f.s since the Stable Laws are, in general, expressed only through their characteristic functions.

A number of researchers have applied stable distributions to practical examples, so this distribution is of more than just academic interest. It has found uses in the modelling of stock price data; see for example Mandelbrot (1963), Fama (1963, 1965) and Koutrouvelis (1980). More recently, it has been employed by Brockwell & Liu (1991) in a Time Series application. They sought to estimate the parameters of the noise model when it was allowed to follow a stable distribution. Remaining in the Time Series setting, the e.c.f. was used by Chan & Tong (1990) to show that an autoregressive model for the lynx data would be inappropriate.

The sequence $\{\hat{\phi}(s):n=1,2,...\}$ has a number of important properties. We begin with a result from Feuerverger and Mureika (1977): for fixed $T<\infty$, we have the convergence

$$Pr\{\lim_{n\to\infty} \sup_{|s|\le T} |\hat{\phi}(s) - \phi(s)| = 0\} = 1. \tag{3.9}$$

Because $\hat{\phi}(s)$ is a trigonometric polynomial, it is essentially periodic and hence must approach its supremum value $\hat{\phi}(0)=1$ arbitrarily often as $|s| \to \infty$; as for example when the distribution function $F(x)$ is absolutely continuous. This uniform convergence cannot therefore be extended to the whole real line in general, although it can be when $F(x)$ is purely discrete. In this case :

$$Pr\{\lim_{n\to\infty} \sup_{|s|<\infty} |\hat{\phi}(s) - \phi(s)| = 0\} = 1. \tag{3.10}$$

Proof

By the Strong Law of Large Numbers, if X is a discrete random variable taking values $\{x_k\}$

with probabilities $\{p_k\}$, then in obvious notation

$$|\hat{\phi}(s) - \phi(s)| = |\sum_k (\hat{p}_k - p_k) e^{isx_k}|$$
$$\leq \sum_k |\hat{p}_k - p_k|$$
$$\rightarrow 0 \ (a.s.).$$

□

Feuerverger and Mureika note after this proof that "The connection between the e.c.f. and kernel-type estimators of probability density functions makes it possible to apply results in density estimation to the e.c.f.". The link between ISE and density estimation was demonstrated via Parseval's Theorem (see Appendix I) by Heathcote (1977). He suggested that the smoothing implicit in density estimation is responsible for some of ISE's desirable properties. He gives in particular an example where no global maximum likelihood estimator is possible, although a consistent local one does exist, but the ISE-estimator is well-defined. We argue here that this a result of employing the e.c.f.. The influence function approach, as we see later, enables us to obtain an explicit interpretation for λ beyond that which is possible via the density estimation route.

Returning to consideration of the sequence $\{\hat{\phi}(s)\}$, we may form

$$Y_n(s) = n^{1/2}\{\hat{\phi}(s) - \phi(s)\} \tag{3.11}$$

to be considered a random complex process in s. This was discussed in Feuerverger and McDunnough (1981), employing the multidimensional central limit theorem to show that $\{Y_n(s_1),...,Y_n(s_r)\}$ converges to $\{Y(s_1),...,Y(s_r)\}$ given that $E(Y_n(s))=0$ and $E(Y_n(s)Y_n(t))=\phi(s+t)-\phi(s)\phi(t)$ exists.

The results depicted here demonstrate that the e.c.f. is a well-behaved function, and likely to be of considerable statistical use.

3.2. Robustness of ISE Methods.

The ISE function is defined to be

$$h(\Theta) = \int_{-\infty}^{\infty} |\hat{\phi}(s) - \phi(s;\Theta)|^2 w(s;\lambda) ds \ , \tag{3.12}$$

where $\phi(.)$ denotes the characteristic function. With regards estimation, we note that the ISE-estimator $\hat{\Theta}$ is such that

$$h(\hat{\Theta}) = \min_{\Theta} h(\Theta)$$

which is found in practice as the solution of

$$\frac{\partial h(\Theta)}{\partial \theta_j} = 0 \quad j=1,...,p. \tag{3.13}$$

Heathcote (1977) shows that, in the regular case allowing for interchange of integration and differentiation, there exists a strongly consistent solution to the system (3.13). Under these mild regularity conditions, equations (3.13) may be written as

$$\int_{-\infty}^{\infty} [(\hat{R}-R)\frac{\partial R}{\partial \theta_j} + (\hat{I}-I)\frac{\partial I}{\partial \theta_j}] w(s;\lambda) ds = 0 \quad j=1,...,p, \tag{3.14}$$

where $R=Re\phi(s;\Theta)$ and $I=Im\phi(s;\Theta)$. Note that equations (3.14) follow readily on rewriting (3.12) as

$$h(\Theta) = \int_{-\infty}^{\infty} \{(\hat{R}-R)^2 + (\hat{I}-I)^2\} w(s;\lambda) ds \ .$$

We go on now to investigate the robustness properties of $\hat{\Theta}$. The following example begins this process with an intuitive discussion of ISE's resistance to outlying values.

Example 3.1.

Suppose that $X_1, \ldots, X_n$ are independent $N(0,1)$ random variables. Let $w(s)=1$, then we may write the ISE function in terms of the data as

$$h(\underline{X}) = \pi^{1/2}[1 - n^{-1}2^{3/2}\sum_{j=1}^{n} e^{-X_j^2/2}]$$

in obvious notation. Now, if we leave-out the ith observation then $h(.)$ becomes

$$h(\underline{X}_{(-i)}) = \pi^{1/2}[1 - (n-1)^{-1}2^{3/2}\sum_{j\neq i} e^{-X_i^2/2}]$$

$$\Rightarrow h(\underline{X}_{(-i)}) - h(\underline{X}) \approx n^{-1}(8\pi)^{1/2}[\, e^{-X_i^2/2} - (n-1)^{-1} \sum_{j \neq i} e^{-X_j^2/2} \,] .$$

We see that the effect on the ISE surface of leaving out the ith observation is bounded. This suggests an intrinsic robustness of the integrated-squared-error technique to outlying observations.

We can make this intuitive treatment precise via the following

Theorem 3.1.

Under the same regularity conditions as Heatcote (1977), the ISE-estimator $\hat{\Theta}$ of Θ has influence function

$$IF(X;\hat{\Theta}) = \kappa^{-1}\tau(X) ,$$

where

$$(\kappa)_{ij} = \int_{-\infty}^{\infty} [\frac{\partial R}{\partial \theta_j}\frac{\partial R}{\partial \theta_i} + \frac{\partial I}{\partial \theta_j}\frac{\partial I}{\partial \theta_i}] w(s;\lambda) ds \qquad i,j=1,...,p$$

and

$$[\tau(X)]_j = \int_{-\infty}^{\infty} [\frac{\partial R}{\partial \theta_j} IF(X;\hat{R}) + \frac{\partial I}{\partial \theta_j} IF(X;\hat{I})] w(s;\lambda) ds \qquad j=1,...,p.$$

Note that $IF(X;\hat{R})$ and $IF(X;\hat{I})$ are as defined by equations (3.8).

□

We see that $\hat{\Theta}$ possesses bounded influence provided R and I are both uniformly differentiable w.r.t. to each of the $\theta_j s$, $j=1,...,p$. As we shall see in an example to come, the sensitivity of the technique is controlled by λ.

The proof of this theorem first requires the following

Lemma

Suppose that a statistic is defined by

$$\hat{Y} = \int T(\hat{F};s)ds \ .$$

Then, by definition

$$\begin{aligned} IF(X;\hat{Y}) &= \frac{\partial}{\partial\varepsilon}[Y\{(1-\varepsilon)F+\varepsilon A\}] \\ &= \frac{\partial}{\partial\varepsilon}\int T\{(1-\varepsilon)F+\varepsilon A\}ds \\ &= \int \frac{\partial}{\partial\varepsilon} T\{(1-\varepsilon)F+\varepsilon A;\, s\}ds \end{aligned}$$

so that

$$IF(X;\hat{Y}) = \int IF(X;\hat{T}(s))ds \ .$$

□

Applying this lemma to (3.14), we find the jth influence equation

$$\int_{-\infty}^{\infty} [\frac{\partial R}{\partial\theta_j}\{IF(X;\hat{R}) - \sum_{i=1}^{P}\frac{\partial R}{\partial\theta_i} IF(X;\hat{\theta}_i)\} + \frac{\partial I}{\partial\theta_j}\{IF(X;\hat{I}) - \sum_{i=1}^{P}\frac{\partial I}{\partial\theta_i} IF(X;\hat{\theta}_i)\}]w(s;\lambda)ds = 0 \qquad j=1,...,p.$$

This provides a vivid demonstration of the flexibility of the influence tools developed in Chapter 2. Now, rearranging the above equations we find

$$\sum_{i=1}^{P} IF(X;\hat{\theta}_i)[\int_{-\infty}^{\infty}(\frac{\partial R}{\partial\theta_j}\frac{\partial R}{\partial\theta_i} + \frac{\partial I}{\partial\theta_j}\frac{\partial I}{\partial\theta_i})w(s;\lambda)ds] =$$

$$\int_{-\infty}^{\infty}[\frac{\partial R}{\partial\theta_j} IF(X;\hat{R}) + \frac{\partial I}{\partial\theta_j} IF(X;\hat{I})]w(s;\lambda)ds \qquad j=1,...,p.$$

That is,

$$\sum_{i=1}^{P}(\kappa)_{ij} IF(X;\hat{\theta}_i) = [\tau(X)]_j \qquad j=1,...,p.$$

□

This result may not provide a particularly efficient method for calculating influence functions. However, it does establish why ISE methods, based on characteristic functions, are robust. The key, once again, is the choice of transform.

Corollary

We note that, by standard theory,

$$var(\hat{\Theta}) \approx n^{-1} \int_{-\infty}^{\infty} IF(X;\hat{\Theta})IF(X;\hat{\Theta})'dF(X)$$

$$= n^{-1} \int_{-\infty}^{\infty} \kappa^{-1}\tau(X)\tau(X)'(\kappa^{-1})'dF(X) .$$

We see from its definition that κ is symmetric, so

$$var(\hat{\Theta}) \approx n^{-1}\kappa^{-1}\Sigma\kappa^{-1} ,$$

where Σ is the dispersion matrix of $\tau(X)$.

This corollary corresponds to a result found by Heathcote (1977), although it follows more readily via the influence function technique. In the following example we illustrate the efficiency and robustness properties of ISE via a particular case.

Example 3.2.

Suppose that we have a random sample from a $N(\mu,\sigma^2)$ distribution available, where σ^2 is assumed known. We choose the weight function $w(s)=e^{-\lambda^2 s^2}$; this Gaussian form is a popular choice used by, for example, Paulson et al. (1975), Heathcote (1977) and Paulson & Nicklin (1983). This popularity seems to be due to two major reasons. First, the Gaussian form is well-known to Statisticians and so likely to be chosen for this application if no decision rule is obvious. Second, it is a valid characteristic function. Thus, applying Parseval's Theorem to the ISE function gives $w(.)$ an interpretation as a smoothing density function.

Recall that

$$\phi(s;\mu,\sigma) = e^{is\mu - s^2\sigma^2/2}$$

so that $R = e^{-s^2\sigma^2/2}\cos(s\mu)$ and $I = e^{-s^2\sigma^2/2}\sin(s\mu)$. Thus

$$\frac{\partial R}{\partial \mu} = -se^{-s^2\sigma^2/2}\sin(s\mu) ,$$

and

$$\frac{\partial I}{\partial \mu} = se^{-s^2\sigma^2/2}\cos(s\mu) .$$

It is straightforward, if rather tedious, to show that

$$\tau = (2\pi)^{1/2}(\sigma^2+2\lambda^2)^{-3/2}(X-\mu)e^{-0.5\,(X-\mu)^2\,/\,(\sigma^2+2\lambda^2)}\;;$$
$$\kappa = 2^{-3/2}(2\pi)^{1/2}(\sigma^2+\lambda^2)^{-3/2}\;,$$

yielding

$$IF\,(X;\hat{\mu}) = 2^{3/2}\left(\frac{\sigma^2+\lambda^2}{\sigma^2+2\lambda^2}\right)^{3/2}(X-\mu)e^{-0.5\,(X-\mu)^2\,/\,(\sigma^2+2\lambda^2)}\;,$$

on applying Theorem 3.1. This result may also be obtained from the theory of M-estimation, employing a similar amount of effort. Such methods are discussed in the standard texts Huber (1981) and Hampel et al. (1986).

We see that the ISE-estimator is robust to outlying observations, whereas the usual estimator, the sample mean, is not (Example 2.1). By decreasing the value of λ, we reduce sensitivity to these observations, giving us some control over this procedure. This is depicted graphically, in Figure 3.1 below, for $\mu = 0$ and $\sigma = 1$.

Note that

$$\lim_{\lambda\to\infty} IF\,(X;\hat{\mu}) = X-\mu$$
$$= IF\,(X;\bar{X})\;.$$

Thus, for large λ, $\hat{\mu}$ behaves as $\bar{X}$. The reason for this becomes obvious when we note that the ISE estimating equation is

$$\sum_{j=1}^{n}(X_j-\hat{\mu})\exp[-0.5\,(X_j-\hat{\mu})^2/(\sigma^2+2\lambda^2)] = 0\;,$$

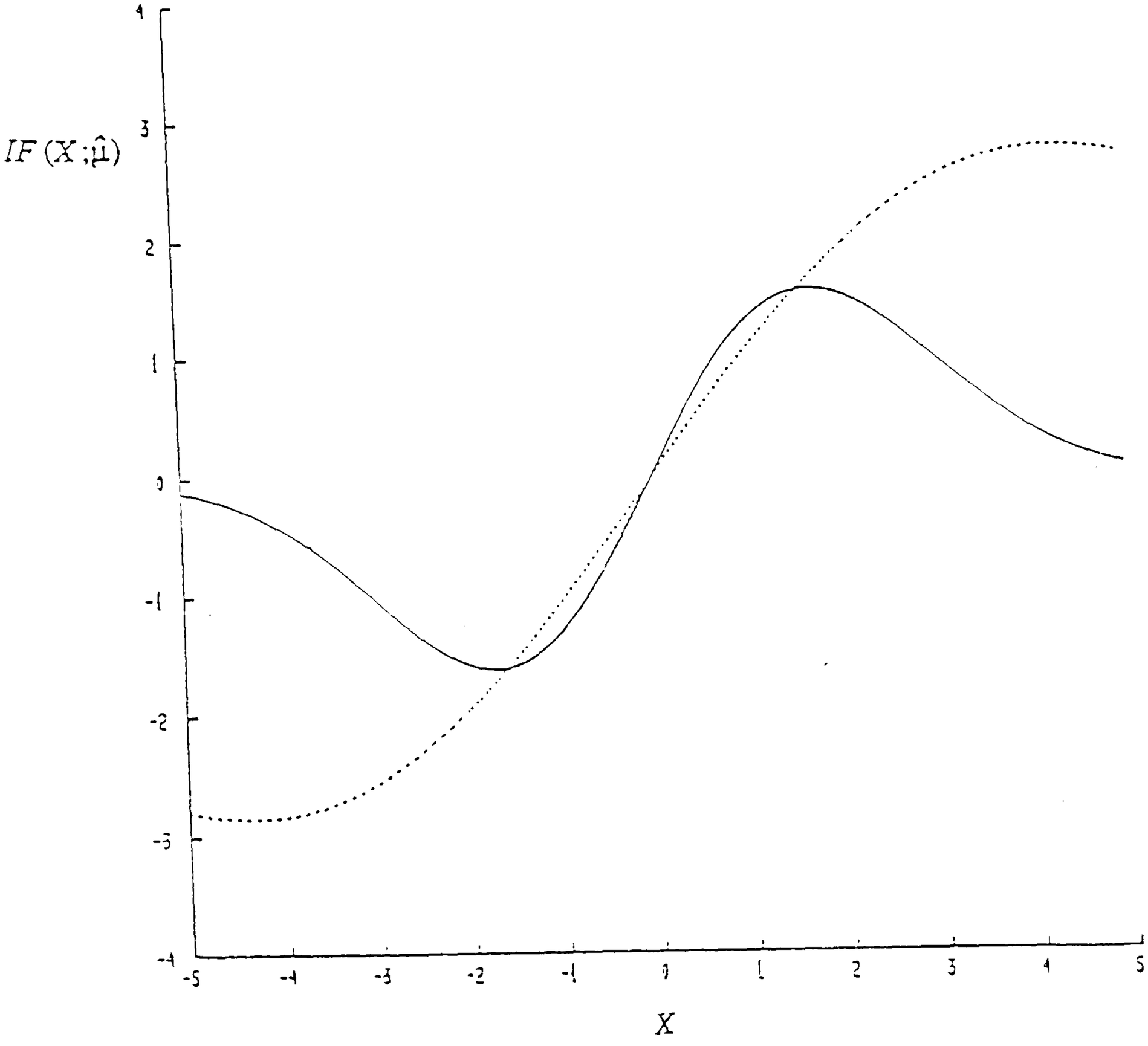
and for large λ this becomes

$$\sum_{j=1}^{n}(X_j-\hat{\mu}) = 0$$
$$\Rightarrow \hat{\mu} = \bar{X}\;.$$

We therefore have the concrete interpretation for λ that it controls the robustness of this procedure. To obtain maximum robustness, we choose $\lambda = 0$. For less protection against outliers, we allow λ to become larger. An alternative interpretation, in terms of density estimation, is discussed in the next section.

Figure 3.1. Influence function for ISE-estimator of normal mean;

solid line, $\lambda = 1$; dotted line, $\lambda = 3$.



Turning now to matters of efficiency, assuming uncontaminated data, we find the asymptotic efficiency of $\hat{\mu}$ versus $\bar{X}$ by applying Result 2.2. In the particular case $\mu = 0$, $\sigma = 1$ this is

$$\frac{1}{8}(1+2\lambda^2)^{3/2}(3+2\lambda^2)^{3/2}(1+\lambda^2)^{-3} .$$

Even when the data are uncontaminated we obtain 90% asymptotic efficiency when $\lambda = 1$, a value providing substantial protection against outliers. When $\lambda = 5$, $\hat{\mu}$ attains 99.9% efficiency but provides an estimator lacking in robustness. The trade-off between efficiency and robustness is a recurring theme of this thesis.

Taking these facts concerning robustness and efficiency together, ISE seems to provide an ideal robust estimator in this case. At the true model, it can attain very reasonable efficiency whilst also being robust to outlying observations. Note that this robustness is very strong here, since $\hat{\mu}$ is IBR with $M = 0$. Away from the true model, as we shall see in Example 3.3, $\hat{\mu}$ can become more efficient than $\bar{X}$. This is what a robust estimator must achieve to be of practical use.

However, given uncontaminated data, $\hat{\mu}$ is necessarily inferior to $\bar{X}$ as an estimator for μ. To complete the picture, we investigate the behaviour of $\hat{\mu}$ versus $\bar{X}$ in the presence of contaminated data via the following

Example 3.3.

Rather than observing, as in the previous example, $N(\mu,\sigma^2)$ random variables we assume that the data are in fact subject to contamination. We model this via the mixture distribution

$$F_X = (1-\varepsilon)\,N(\mu,\sigma^2)+\varepsilon\,N(\mu,3\sigma^2) ,$$

in obvious notation. A contamination model of this sort was suggested by Huber (1981) pp 2-3 as being physically reasonable.

We compare the performance of $\hat{\mu}$ versus $\bar{X}$ via the variance ratio

$$\Omega = var(\bar{X})/var(\hat{\mu}) .$$

Applying Result 2.2, we find after a little algebra that

$$var(\hat{\mu}) \approx n^{-1}\sigma^2(\lambda^2+\sigma^2)^3(\lambda^2+\sigma^2/2)^{-3/2}\,[(1-\varepsilon)(\lambda^2+3/2\sigma^2)^{-3/2}+3\varepsilon\,(\lambda^2+7/2\sigma^2)^{-3/2}] ,$$

so that

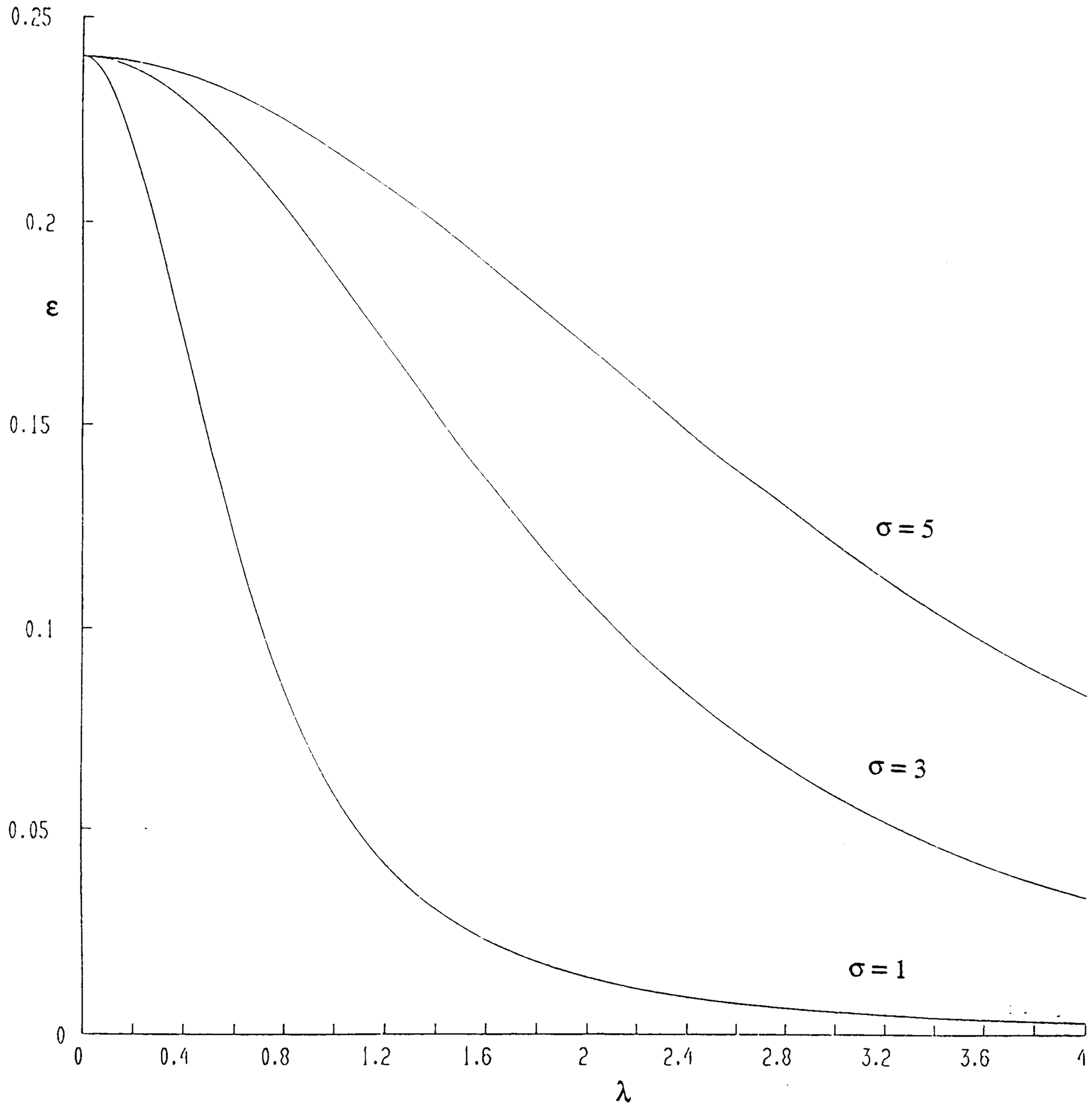
$$\Omega = \frac{(1+3\varepsilon)(\lambda^2+\sigma^2/2)^{3/2}}{(\lambda^2+\sigma^2)^3[(1-\varepsilon)(\lambda^2+3/2\sigma^2)^{-3/2}+3\varepsilon\,(\lambda^2+7/2\sigma^2)^{-3/2}]}\,.$$

Figure 3.2 below is a plot of ε against λ for three values of σ on the boundary $\Omega = 1$.

For large contamination, we see that a small value for λ is required to achieve $\Omega = 1$, corresponding to a more robust estimator being preferred in this situation. As the level of contamination falls, we see that λ increases and $\hat{\mu}$ behaves more like $\bar{X}$. This clearly accords well with common sense. We concentrate now on a particular ISE-estimator, that for $\lambda = 2$ say. As σ increases, we see that the level of contamination required to maintain $\Omega = 1$ tends to increase. This is because the data are naturally more noisy, so small to moderate contamination is difficult to detect. Note that for any particular comparison of $\hat{\mu}$ versus $\bar{X}$, adding more contamination will tend to tip the balance in favour of $\hat{\mu}$.

The work so far on ISE has first established the source of its apparent robustness, which is fundamentally because estimation is based on the e.c.f. Secondly, we have seen that ISE can compete well with standard methods not only in terms of robustness but efficiency as well. We have made clear the *rôle* played by λ in particular. In the next section we offer a further interpretation for this parameter.

Figure 3.2. Plot of ε against λ on the boundary $\Omega = 1$.



3.3. An Alternative Interpretation for λ.

Returning to the general definition of ISE (3.12), we may write

$$h(\Theta) = \int_{-\infty}^{\infty} |\hat{\psi}(s;\lambda) - \psi(s;\Theta,\lambda)|^2 ds \ , \tag{3.15}$$

where $\psi(s;\Theta,\lambda) = \phi(s;\Theta)[w(s;\lambda)]^{1/2}$. Provided $[w(s;\lambda)]^{1/2}$ is itself a valid characteristic function, then so is $\psi(.)$. This is certainly the case for $w(s;\lambda)=e^{-\lambda^2 s^2}$, and we assume this to be true for the general $w(.)$ we employ in this section.

Applying Parseval's Theorem (Appendix I) to (3.15), we find

$$h(\Theta) = 2\pi \int_{-\infty}^{\infty} [\hat{g}(x;\lambda) - g(x;\Theta,\lambda)]^2 dx \ , \tag{3.16}$$

where $g(x;\Theta,\lambda)$ has characteristic function $\psi(s;\Theta,\lambda)$, and similarly for $\hat{g}(x)$. Thus

$$\begin{aligned}\hat{g}(x) &= (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-isx} \hat{\psi}(s;\lambda) ds \\ &= n^{-1} \sum_{j=1}^{n} (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-isx} e^{isX_j - 1/2 \log w(s;\lambda)} ds \ .\end{aligned}$$

We focus now on $w(s;\lambda)=e^{-\lambda^2 s^2}$, so that

$$\begin{aligned}\hat{g}(x) &= n^{-1} \sum_{j=1}^{n} (2\pi)^{-1} \int_{-\infty}^{\infty} e^{-isx} e^{isX_j - \lambda^2 s^2/2} ds \\ &= n^{-1} \sum_{j=1}^{n} (2\pi\lambda^2)^{-1/2} e^{-(x-X_j)^2/(2\lambda^2)} \ .\end{aligned} \tag{3.17}$$

Thus we see that λ has a *rôle* as a bandwidth in the density estimation problem (3.16). Large λ (lacking robustness) corresponds to large smoothing. For small λ (more robust), less smoothing takes place.

This interpretation for λ has been noted by a number of researchers, including Heathcote (1978) and Bryant & Paulson (1983).

3.4. Choosing a Value for λ.

We mention a few ideas here for completeness only. This topic is not central to the aims of this thesis, so we do not pursue them in detail.

3.4.1. Asymptotic Efficiency.

Here we choose λ so that asymptotic efficiency exceeds some critical value at the true model. A value of 90% seems a reasonable guide, yielding a choice of $\lambda = 1$ in Example 3.2.

3.4.2. Numerical Choice.

One could simply include λ as an extra parameter to be estimated, viz:

$$h(\hat{\Theta},\hat{\lambda}) = \min_{(\Theta,\lambda)} h(\Theta,\lambda).$$

This method seems rather unappealing in that it adds an extra dimension to what may already be a high-dimensional search. In addition, we see that λ and σ are not identifiable in Example 3.2 where we can only estimate the pair $(\mu,\eta = \sigma^2 + 2\lambda^2)'$. This is a consequence of the weight-function being the characteristic function of a Normal distribution, so λ becomes incorporated in the dispersion parameter of the model. We observe a *rôle* for the weight function similar to that of a conjugate prior in Bayesian statistics.

A further criticism of this idea is that it tries to " get something for nothing ". The data are used to estimate Θ and also to select the robustness of the procedure. Overall, it seems that an external criterion for selecting λ is appropriate.

3.4.3. Using Parameters of the IF.

It was suggested in Bryant & Paulson (1983) that λ be chosen from considerations of robustness. We discuss here choosing λ to minimize the Gross-Error Sensitivity. This technique is well illustrated by

Example 3.2 - continued.

For this case, it is straightforward to show that

$$\gamma = 2^{3/2} e^{-1/2} (\lambda^2+\sigma^2)^{3/2} / (2\lambda^2+\sigma^2) ,$$

yielding

$$\begin{aligned} \hat{\lambda} &= 2^{-1/2}\sigma \\ &\approx 0.707\sigma . \end{aligned}$$

This choice of λ gives an efficiency of 84% for $\hat{\mu}$ versus $\bar{X}$, slightly less than that achieved by the

first rule outlined above. This is perhaps unsurprising, since this method places direct emphasis on the robustness of the resulting estimator.

4. The Method-of-Moments.

The simplest technique for model-fitting is the method-of-moments. Suppose that we wish to estimate a vector of unknown parameters, Θ, as defined earlier. For this method we equate the empirical and theoretical transforms at each $s \in S = \{s_1, \ldots, s_p\}$ to obtain the p estimating equations

$$\hat{G}(s_i) = G(s_i; \Theta(\hat{F})) \quad i=1,...,p. \tag{3.18}$$

Note the explicit dependence on the empirical c.d.f. reflected in the notation used. This is important to remember, since $\hat{\Theta}$ is merely a functional of $\hat{F}$, even if rather more complicated than those we have discussed so far.

We may view the method-of-moments, as noted by Heathcote (1977), as a special case of ISE, since if we choose

$$w(s;\lambda) = \delta(s - s_k) \quad s_k \in S \tag{3.19}$$

in (3.12), where $\delta(.)$ represents the kronecker delta function, then

$$h(\Theta) = \sum_{i=1}^{p} |\hat{G}(s_i) - G(s_i; \Theta)|^2 .$$

As we normally employ real-valued transforms in this technique, $\hat{\Theta}$ results from minimizing

$$h(\Theta) = \sum_{i=1}^{p} [\hat{G}(s_i) - G(s_i; \Theta)]^2 ,$$

so that we are required to solve the system (3.18).

4.1. Influence Theory.

Influence functions for moment estimators are found by applying

Theorem 3.2.

Given a well-defined moment estimator $\hat{\Theta}$ of Θ, we have

$$IF(X;\hat{\Theta}) = \nabla^{-1} IF(X;\underline{\hat{G}}),$$

where

$$[IF(X;\underline{\hat{G}})]_i = IF(X;\hat{G}(s_i)) \qquad i=1,...,p;$$
$$[IF(X;\hat{\Theta})]_i = IF(X;\hat{\Theta}_i) \qquad i=1,...,p;$$

$$(\nabla)_{ij} = \frac{\partial G(s_i)}{\partial \theta_j} \qquad i,j=1,...,p.$$

□

Proof

For the ith equation of system (3.18) we may write-down the influence equation

$$IF(X;\hat{G}_i) = \sum_{j=1}^{p} \frac{\partial G(s_i)}{\partial \theta_j(F)} IF(X;\hat{\theta}_j),$$

where we employ the notation $G_i = G(s_i)$ as convenient. Bringing these p influence equations together, we may form the single matrix equation

$$IF(X;\underline{\hat{G}}) = \nabla\, IF(X;\hat{\Theta}).$$

We note that ∇ is a Jacobian matrix, so given the context it seems valid to assume that ∇ is non-singular. Thus,

$$IF(X;\hat{\Theta}) = \nabla^{-1} IF(X;\underline{\hat{G}}).$$

□

It is clear from this that the parameter estimators will inherit the influence properties of the empirical transforms employed. In particular, Θ will be IBR (see page 21) iff G is. Note further that $IF(X;\underline{\hat{G}})$ is an elementary calculation, and so $IF(X;\hat{\Theta})$ is generally easy to calculate for moment estimators, even if no explicit solution for $\hat{\Theta}$ exists. The calculation of ∇ is especially easy now that symbolic algebra packages are readily available.

Example 3.4.

Suppose that a random sample of n observations is available from the $N(\mu,\sigma^2)$ distribution. We base estimation on the moment generating function in this example, so that the estimating equations are

$$\hat{M}(s_i) = \exp(s_i\hat{\mu} + s_i^2\sigma^2/2) \quad i=1,2.$$

As a first step, we obtain the first row of ∇ since, by symmetry, we can then fill-in the rest of the matrix to obtain

$$\nabla = \begin{bmatrix} s_1 M(s_1) & s_1^2/2\, M(s_1) \\ s_2 M(s_2) & s_2^2/2\, M(s_2) \end{bmatrix}.$$

By elementary methods we obtain, in obvious notation,

$$IF(X;\underline{\hat{M}}) = (e^{s_1 X} - M(s_1), e^{s_2 X} - M(s_2))',$$

so that after inverting ∇ we obtain, on writing $\Theta = (\mu,\sigma^2)'$

$$IF(X;\hat{\Theta}) = (s_2 - s_1)^{-1} \begin{bmatrix} \frac{s_2}{s_1 M_1}(e^{s_1 X} - M_1) - \frac{s_1}{s_2 M_2}(e^{s_2 X} - M_2) \\ \frac{2}{s_2 M_2}(e^{s_2 X} - M_2) - \frac{2}{s_1 M_1}(e^{s_1 X} - M_1) \end{bmatrix};$$

it is straightforward to check that $E\,[IF(X;\hat{\Theta})] = \underline{0}$.

It is instructive to verify this result by only allowing ourselves access to elementary methods. The first step is to solve the estimating equations for $\hat{\mu}$ and $\hat{\sigma^2}$. We can do this explicitly here, but this will not always be the case. Thus,

$$s_1 s_2 (s_2 - s_1)\hat{\mu} = s_2^2 \log \hat{M}_1 - s_1^2 \log \hat{M}_2$$

and

$$s_1^2 \hat{\sigma^2} = 2\,[\log \hat{M}_1 - s_1\hat{\mu}],$$

yielding

$$s_1 s_2 (s_2 - s_1) IF(X;\hat{\mu}) = \frac{s_2^2}{M_1}(e^{s_1 X} - M_1) - \frac{s_1^2}{M_2}(e^{s_2 X} - M_2)$$

and

$$s_1^2/2\, IF(X;\hat{\sigma^2}) = M_1^{-1} IF(X;\hat{G}_1) - s_1 IF(X;\hat{\mu})$$

giving the earlier result when simplified.

This example was really very easy to verify. This is not so easy to do in the following case which demonstrates the usefulness of these methods.

Example 3.5.

We look again at the Negative-Binomial distribution as parameterized in Example 2.5. The estimating equations are

$$\hat{\Pi}(z_i) = [1+(1-z_i)\hat{q}]^{-\hat{k}} \qquad i=1,2.$$

Writing $w_i=1+(1-z_i)q$ and employing the notation $\Pi_i=\Pi(z_i)$, we form

$$\nabla = \begin{bmatrix} -k(1-z_1)w_1^{-1}\Pi_1 & k^{-1}\Pi_1\log\Pi_1 \\ -k(1-z_2)w_2^{-1}\Pi_2 & k^{-1}\Pi_2\log\Pi_2 \end{bmatrix}.$$

Now,

$$IF(X;\underline{\hat{\Pi}}) = \begin{bmatrix} z_1^X - \Pi_1 \\ z_2^X - \Pi_2 \end{bmatrix}.$$

We find, in particular, that

$$IF(X;\hat{q}) = \frac{\Pi_2[z_1^X-\Pi_1]\log\Pi_2 - \Pi_1[z_2^X-\Pi_2]\log\Pi_1}{k\Pi_1\Pi_2\log[\Pi_1^{-(1-z_2)/w_2}\Pi_2^{-(1-z_1)/w_1}}$$

We have no other means available at the moment to verify this result. Later in this chapter we develop further methods which will allow us to do so.

4.2. Variance Considerations.

Appealing to equation (2.16), we have the definition

$$\begin{aligned} n\,var(\hat{\Theta}) &\approx E_F[\,IF(X;\hat{\Theta})IF(X;\hat{\Theta})'\,] \\ &= E_F[\,\nabla^{-1}IF(X;\underline{\hat{G}})IF(X;\underline{\hat{G}})'(\nabla^{-1})'\,] \\ \Rightarrow var(\hat{\Theta}) &\approx n^{-1}\nabla^{-1}var(\underline{\hat{G}})(\nabla^{-1})'. \end{aligned} \tag{3.20}$$

Thus, given $var(\underline{\hat{G}})$, measures of error for parameter estimators are easy to calculate. On this point, the dispersion matrix of transforms of the form (3.1) is very easy to find, as follows.

$$\begin{aligned} cov(\hat{G}_i,\hat{G}_j) &= n^{-2}cov[\sum_{r=1}^{n} g(s_i,X_r), \sum_{m=1}^{n} g(s_j,X_m)] \\ &= n^{-2}\sum_{r,m=1}^{n} cov[g(s_i,X_r),g(s_j,X_m)] \end{aligned}$$

$$= n^{-2} \sum_{r,m=1}^{n} [E_F(g(s_i,X_r)g(s_j,X_m)) - G_iG_j] \,. \tag{3.21}$$

Now,

$$E_F[g(s_i,X_r)g(s_j,X_m)] = \int g(s_i,X)g(s_j,X)dF(X) \,,$$

given that $X_1, \ldots, X_n$ are identically distributed. The result of this integral depends on the form of the kernel, and can typically be written in terms of G itself. For example, when the transform of interest is the characteristic or moment generating function then the integral is $G(s_i + s_j)$. The probability generating function is rather different, yielding $G(s_i s_j)$. Clearly $var(\underline{\hat{G}})$ is very easy to find for standard transforms.

Returning now to (3.20), we note that this result corresponds to that obtained by standard techniques, since applying the delta method yields

$$var[G(\hat{\Theta})] \approx \sum_{i,j=1}^{p} cov(\hat{\theta}_i,\hat{\theta}_j) \frac{\partial G}{\partial \theta_i} \frac{\partial G}{\partial \theta_j} \,.$$

This is an approximation for $var(\underline{\hat{G}})$, where the influence technique yielded a solution directly for $\hat{\Theta}$.

4.3. Choosing $\underline{s}$.

In this section we consider choosing the optimal value of $\underline{s}$, $\underline{s}^*$ say, by minimizing the variance of resulting estimators. In the general p-parameter setting the criterion we apply is that of generalized variance, denoted by $gv(.)$, so that

$$gv(\underline{s}^*) = \min_{\underline{s}} |var(\hat{\Theta}(\underline{s}))| \tag{3.22}$$

in obvious notation. Recalling (3.23),

$$gv(\underline{s}^*) = \min_{\underline{s}} |\nabla^{-1} var(\underline{\hat{G}})(\nabla^{-1})'|$$

where explicit dependence on $\underline{s}$ has been suppressed for ease of exposition. Thus,

$$gv(\underline{s}^*) = \min_{\underline{s}} |var(\underline{\hat{G}})| \, |\nabla|^{-2} \,. \tag{3.23}$$

We see that $\underline{s}^*$ can be selected even if $\hat{\Theta}$ is not known explicitly. This method will be used extensively in Chapter 4.

One of the problems with this technique is that the objective function of (3.25) depends on the unknown parameters Θ. We can overcome this difficulty by substituting $\hat{\Theta}$ for Θ, much as an estimated standard error is used in the standard t-test when the population variance is unknown. We investigate this in an example to come.

We take the opportunity here to note an interesting phenomenon of this work, which is best illustrated by

Example 3.6.

Suppose that $X_1, \ldots, X_n$ is a random sample from the exponential distribution having mean λ^{-1}. Then

$$\phi(s;\lambda) = (1-is/\lambda)^{-1} .$$

Using the method of moments, based on the ratio of real and imaginary parts, we find

$$\hat{\lambda}(s) = s\sum_{j=1}^{n}\cos(sX_j)/\sum_{j=1}^{n}\sin(sX_j)$$

and subsequently by the method of influence functions that

$$var(\hat{\lambda}) \approx n^{-1}(1+s^2)^2(1+2s^2)(1+4s^2)^{-1} ,$$

This suggests that we take $s^* = 0$. We see that the reason for this is because

$$\lim_{s\to 0}\hat{\lambda}(s) = \bar{X}^{-1} ,$$

the m.l.e. for λ, which attains the Cramér-Rao lower bound. Thus we can only make a degenerate choice of s in this case.

Based on examples, this seems to happen only for one-parameter cases where the m.l.e. results as a moment estimator, and so includes all members of the exponential family. We can justify this by the following semi-intuitive argument. The estimator concerned results as the solution of

$$\hat{\phi}(s) = \phi(s;\hat{\theta})$$

so expanding the characteristic functions we obtain, for complex k,

$$1+is\bar{X}+(is)^2\hat{k} = 1+is\mu(\hat{\theta})+(is)^2k(\hat{\theta})$$

$$\Rightarrow \bar{X}-\mu(\hat{\theta})+is(\hat{k}-k(\hat{\theta})) = 0$$

and letting $s \rightarrow 0$,

$$\bar{X} = \mu(\hat{\theta}) .$$

So we find that $\hat{\theta}$ arrives as the solution of the first moment equation. In many cases this coincides with the m.l.e. of θ, so that variance-based methods become degenerate.

We note here the possibility of this extending, in a very special case, to the general p-parameter setting. If $cov(\hat{\theta}_i, \hat{\theta}_j) = 0 \ \forall i \neq j$, then

$$| var(\underline{\hat{\Theta}}) | = \prod_{i=1}^{p} var(\hat{\theta}_i(\underline{s})) .$$

This suggests that we can minimize the generalized variance by choosing $\underline{s} = \underline{0}$. However, this is unlikely for at least two reasons. First, it is rare for all parameter estimators to be uncorrelated. Second, the earlier argument suggests that this phenomenon only extends as far as the first moment; higher-order moments appear to be exempt. This is a problem which may benefit from further theoretical work, but is somewhat outside the remit of this thesis.

5. Extensions to Influence Theory.

In this section we extend the influence theory we have developed in two distinct ways. The first is an extension of the method-of-moments to transforms which cannot be easily written in the form (3.1). Secondly, we demonstrate the range of application of our methods by applying them to really quite general estimating equations.

5.1. Non-Standard Transforms.

Consider the general set of estimating equations

$$\hat{\psi}_i = \psi_i(s\,;\, \Theta(\hat{F})) \qquad i=1,\ldots,p. \tag{3.24}$$

One choice of ψ, justification for which is provided in Appendix II, is

$$\psi_i = \psi^{(i-1)} \qquad i=1,\ldots,p,$$

with the superscript $(i-1)$ denoting the $(i-1)$th derivative w.r.t. s. Note that we take the case $i=1$ to mean the function ψ itself.

Clearly the earlier work applies, and we have the result

$$IF(X;\hat{\Theta}) = \nabla^{-1} IF(X;\underline{\hat{\psi}}), \tag{3.25}$$

where

$$(\nabla)_{ij} = \frac{\partial \psi_i(s)}{\partial \theta_j} \qquad i,j=1,\ldots,p$$

and

$$[\, IF(X;\underline{\hat{\psi}}\,]_i = IF(X;\hat{\psi}_i) \qquad i=1,\ldots,p. \tag{3.26}$$

It is this vector of influence functions where the real difference occurs. To calculate the elements we typically have to apply a little more effort than for standard transforms, as we shall see in Chapter 4. The following example provides a problem which fits into this general framework, but where the influence vector is easy to calculate. We take this opportunity to bring together all the strands we have discussed thus far and, in addition, to make a comparison in terms of both robustness and efficiency with maximum likelihood. As such, this is a very important example.

Example 3.7.

A number of researchers have used transforms to fit the stable laws. For example, Paulson et al. (1975) and Koutrouvelis (1981). Koutrouvelis (1982) considers the Cauchy distribution in particular, which we discuss here.

Given a random sample $X_1, \ldots, X_n$ from such a distribution having location α and scale β, the characteristic function is

$$\phi(s;\Theta) = e^{is\alpha - \beta|s|}, \quad \Theta = (\alpha,\beta)'.$$

We may form the estimating equations as

$$\begin{aligned} \text{Re}\,\hat{\phi}(s) &= \text{Re}\,\phi(s;\hat{\Theta}) \\ \text{Im}\,\hat{\phi}(s) &= \text{Im}\,\phi(s;\hat{\Theta}). \end{aligned}$$

In our notation above, then

$$\psi_1(s) = \text{Re}\,\phi(s) \quad \text{and} \quad \psi_2(s) = \text{Im}\,\phi(s).$$

Explicitly, the estimating equations are

$$n^{-1}\sum_{j=1}^{n}\cos(sX_j) = e^{-\beta|s|}\cos(s\hat{\alpha})$$

$$n^{-1}\sum_{j=1}^{n}\sin(sX_j) = e^{-\beta|s|}\sin(s\hat{\alpha})\,.$$

Thus,

$$\nabla = \begin{bmatrix} -se^{-\beta|s|}\sin(s\alpha) & -|s|e^{-\beta|s|}\cos(s\alpha) \\ se^{-\beta|s|}\cos(s\alpha) & -|s|e^{-\beta|s|}\sin(s\alpha) \end{bmatrix}.$$

Note that the vector of influence functions for the transforms is very easy to calculate because the characteristic function is a complex sum of the sine and cosine transforms of F. As such, we immediately have two distinct transforms of the form (3.1) available. We find, therefore, that

$$IF(X;\hat{\Theta}) = \begin{bmatrix} s^{-1}e^{\beta|s|}\sin s(X-\alpha) \\ |s|^{-1}[1-e^{\beta|s|}\cos s(X-\alpha)] \end{bmatrix}.$$

We see from this that the influence functions oscillate infinitely, so that they do not decay as $|X-\alpha|$ becomes large. It is instructive to compare this behaviour with that of the m.l.e.s of α and β. Now, the m.le.s arise from the estimating equations

$$\sum_{i=1}^{n}\frac{(X_i-\hat{\alpha})}{\hat{\beta}^2+(X_i-\hat{\alpha})^2} = 0$$

$$n\beta^{-1} - \sum_{i=1}^{n}\frac{2\hat{\beta}}{\hat{\beta}^2+(X_i-\hat{\alpha})^2} = 0\,.$$

Appealing to the theory of M-estimation, we find

$$IF(X;\hat{\alpha}) \propto \frac{(X-\alpha)}{(X-\alpha)^2+\beta^2}$$

and

$$IF(X;\hat{\beta}) \propto \frac{2\beta}{(X-\alpha)^2+\beta^2} - \beta^{-1}\,.$$

It is clear from this that the m.l.e.s follow a rather different pattern of influence behaviour. An additional, attractive, feature is that they decay as $|X-\alpha|$ becomes large. Both estimators are IBR, although in the case of $\hat{\alpha}$ we also have $M = 0$ (see page 21).

Turning now to variance considerations, it follows readily, on writing $\underline{\Psi} = [\mathrm{Re}\ \phi(s), \mathrm{Im}\ \phi(s)]'$, that

$$var(\underline{\hat{\Psi}}) = n^{-1}2^{-1}[1-e^{-2\beta|s|}]I_2 ,$$

yielding

$$var(\hat{\alpha},\hat{\beta}) \approx n^{-1}2^{-1}s^{-2}[e^{2\beta|s|}-1]\,I_2 ,$$

where I_2 denotes the identity matrix. The same result may be obtained by, for example, the delta method, although it follows far more readily by the technique employed here.

We now move on to the question of choosing s by minimizing the generalized variance of the estimators. Doing this, we find $\beta|s| = 0.7968$ at the optimum. Using this result, we can now calculate the asymptotic efficiency of the method-of-moments for this example. Note first that the information matrix is $diag\,[n/(2\beta^2), n/(2\beta^2)]$, so lower bounds for $var(\hat{\alpha})$ and $var(\hat{\beta})$ follow readily. Thus, we find the efficiency for $\hat{\alpha}$ and $\hat{\beta}$ to be 65%, the same as for ISE in Example 3.2 when $\lambda = 0$. This value of λ corresponds to maximum robustness, and we had the flexibility to improve efficiency by increasing the value of λ. Feuerverger & McDunnough (1981) obtain an efficiency of 61% for their procedure. They do, however, allow for evaluation of the transform at more than one value of s, so that arbitrarily high efficiency can be attained. Bearing in mind that maximum likelihood produces IBR estimators, one could argue that in this case the cost of obtaining explicit estimators under the transform approach is too high.

To conclude this example we include a note on data-based choice of s^*. Here we calculate

$$gv(s^*) = \min_s s^{-2}(e^{2\hat{\beta}|s|}-1)$$

and substituting the expression for $\hat{\beta}$ we obtain a purely data-based method of selecting s^*. Examples suggest that this objective function can be badly behaved, even for moderately large samples.

Morgan (1990, pers. comm.) suggested an iterative approach. Given an initial estimate, β_0 say, we calculate

$$gv(s_1) = \min_s s^{-2}[e^{2\beta_0|s|}-1]$$

yielding $\beta_0|s_1| = 0.7968$. We then calculate $\beta_1 = \hat{\beta}(s_1)$, repeating the process until convergence.

A robust starting value to take might be to choose β_0 as the semi-interquartile range.

Restricting our attention to $s > 0$, this scheme may be summarized as :

(1) $\beta_r \, s_{r+1} = 0.7968 \quad r = 0,1,2,...$

(2) $\beta_{r+1} = \hat{\beta}\,(s_{r+1})$,

repeat until convergence.

In examples we observe a range of performance. For β as large as 1, convergence did not occur. For smaller β, convergence did occur but not uniformly. On occasion, oscillation between solutions was observed. It seems that selection of s for this example is a non-trivial problem.

5.2. Estimating Equations.

The previous generalization showed how we can use transforms other than the most straightforward variety in the context of moment estimation. We can broaden our study to include a discussion of influence theory for estimators resulting from general estimating equations. Such estimators have been the subject of considerable work recently. See, for example, McCullagh & Nelder (1989) pp 339-352 and Crowder (1986, 87) and references contained therein.

Suppose that we have the system of estimating equations

$$h_i(\,\Theta(\hat{F}),\, \underline{T}_i(\hat{F})\,) = 0 \quad i = 1,...,p. \tag{3.27}$$

Here $\underline{T}_i(\hat{F}) = (T_{i\,1}(\hat{F}),...,T_{it_i}(\hat{F}))'$ is a vector of sample statistics with t_i denoting the number of such statistics present in the *ith* estimating equation. This general framework includes the following special cases :

(i) Let L denote a log-likelihood. The maximum likelihood estimating equations are

$$\frac{\partial L(\Theta)}{\partial \theta_i} = 0 \quad i = 1,...,p.$$

We see that

$$h_i = \frac{\partial L}{\partial \theta_i} \, .$$

(ii) Moment estimation fits into this framework :

$$h_i = \hat{G}(s_i) - G(s_i;\Theta(\hat{F})) \, .$$

The following theorem allows us to calculate influence functions for well-defined estimators resulting from system (3.27).

Theorem 3.3.

$$IF\,(X\,;\,\hat{\Theta}\,) = -\,\nabla_h^{-1} IF\,(X:\underline{\hat{T}}) \, ,$$

where

$$[\,IF\,(X\,;\,\hat{\Theta}\,)\,]_i = IF\,(X:\hat{\theta}_i) \qquad\qquad i=1,...,p,$$

$$[\,IF\,(X\,;\,\underline{\hat{T}}\,)\,]_i = \sum_{l=1}^{t_i} \frac{\partial h_i}{\partial T_{il}} IF\,(X\,;\hat{T}_{il}\,) \qquad i=1,...,p,$$

$$(\nabla_h)_{il} = \frac{\partial h_i}{\partial \theta_l} \qquad i,l=1,...,p.$$

□

Proof

We can write-down, from (3.27), the influence identity

$$\sum_{j=1}^{p} \frac{\partial h_i}{\partial \theta_j} IF\,(X\,;\hat{\theta}_j) + \sum_{l=1}^{t_i} \frac{\partial h_i}{\partial T_{il}} IF\,(X\,;\hat{T}_{il}) = 0 \qquad i=1,...,p.$$

These equations may be brought together into the single matrix equation

$$\nabla_h IF\,(X\,;\hat{\Theta}) + IF\,(X\,;\underline{\hat{T}}) = \underline{0} \, ,$$

and, as before, the assumption of non-singularity seems valid based on the Jacobian form of ∇_h.

□

Example 3.5 - continued.

Define

$$\hat{T} = \log\hat{\Pi}_1 / \log\hat{\Pi}_2 \, ,$$

then the estimating equations yield

$$\hat{T}\,\log[1+(1-z_2)\hat{q}] - \log[1+(1-z_1)\hat{q}] = 0 \, .$$

We have the identity

$$\frac{\partial h}{\partial q} IF(X;\hat{q}) + \frac{\partial h}{\partial T} IF(X;\hat{T}) = 0\,,$$

where

$$h(\hat{q},\hat{T}) = \hat{T}\log[1+(1-z_2)\hat{q}] - \log[1+(1-z_1)\hat{q}]\,.$$

Firstly, by elementary methods

$$IF(X;\hat{T}) = (\Pi_1 \log\Pi_2)^{-1} IF(X;\hat{\Pi}_1) + \Pi_2^{-1}\log\Pi_1 .- (\log\Pi_2)^{-2} IF(X;\hat{\Pi}_2)\,.$$

Further,

$$\frac{\partial h}{\partial q} = w_2^{-1}(1-z_2)T - w_1^{-1}(1-z_1)$$

and

$$\frac{\partial h}{\partial T} = \log w_2\,.$$

Applying the results of this section to these elements yields the same solution as before.

5.3. General Location-Scale Families.

The work of this section was prompted by the Cauchy example (3.7), the results for which can be generalized to some degree.

Suppose that the random variable X, with characteristic function $\phi_X(s)$, has location 0 and scale 1. Let $Y = \mu + \sigma X$, then

$$\phi_Y(s) = e^{is\mu}\phi_X(s\sigma)\,, \tag{3.28}$$

so that Y has location μ and scale σ. The exact nature of this scaling will depend on the form of $\phi_X(.)$.

In notation introduced towards the beginning of this chapter, the moment estimating equations are

$$\hat{R} = \phi_X(s\hat{\sigma})\cos s\hat{\mu}$$
$$\hat{I} = \phi_X(s\hat{\sigma})\sin s\hat{\mu}.$$

Thus,

$$\hat{\mu}(s) = s^{-1}\tan^{-1}(\hat{I}/\hat{R})\,, \tag{3.29}$$

whilst the estimate of $\sigma(s)$ is obtained from

$$\phi_X(s\hat{\sigma}) = n^{-1}\hat{R}/\cos s\hat{\mu}\,. \tag{3.30}$$

It follows readily that

$$IF(X;\hat{\mu}) = \frac{\sin s(X-\mu)}{s\phi_X(s\sigma)}\,, \tag{3.31}$$

so we can find the asymptotic variance of $\hat{\mu}(s)$ in the usual way as

$$n\ var(\hat{\mu}(s)) \approx \int_{-\infty}^{\infty} \frac{\sin^2 s(X-\mu)}{[s\phi_X(s\sigma)]^2}\, dF(X-\mu)$$

$$= \frac{2^{-1}}{[s\phi_X(s\sigma)]^2} \int_{-\infty}^{\infty} [1-\cos(2s\sigma y)]dF(y)$$

$$\Rightarrow n\ var(\hat{\mu}(s)) \approx \frac{1-\mathrm{Re}\,\phi(2s\sigma)}{2[s\phi_X(s\sigma)]^2}\,. \tag{3.32}$$

Example 3.7 - continued.

For the Cauchy,

$$\phi_X(s) = e^{-|s|}\,,$$

and applying (3.32) we find

$$n\ var(\hat{\mu}(s)) \approx \frac{1-e^{-2\sigma|s|}}{2s^2e^{-2\sigma|s|}}$$

$$= \frac{e^{2\sigma|s|}-1}{2s^2}\,,$$

corresponding to the result previously obtained.

We don't investigate results for $\hat{\sigma}(s)$ here, since these depend on the existence and properties of the inverse $\phi_X^{-1}(.)$. Provided it does exist, similar results to those obtained above will follow.

CHAPTER 4:

APPLICATION OF TECHNIQUES TO THE LAGGED-NORMAL DISTRIBUTION.

1. Introduction.

In Chapter 3 we developed a flexible and powerful theory of influence functions, applied in the transform setting. We saw that influence functions do not only convey information regarding robustness, but also provide an often convenient method for calculating asymptotic variances, and are finding increasing use for this purpose. See, for example, Benichou & Gail (1990). In the context of moment estimation, modern symbolic algebra packages, such as MAPLE (see: Maple User Guide (1989)), allow rapid calculation of the Jacobian matrix ∇. As a result, influence behaviour and asymptotic variances for estimators are readily obtained.

In this chapter we concentrate on a particular example, emphasizing the use of symbolic algebra as an aid to this work. The example we choose is the Lagged-Normal distribution; the reasons for this choice are three-fold. First, the Lagged-Normal density function has found important practical applications as a regression model for indicator-dilution curves, as discussed by Davis & Kutner (1976). Second, the density does not exist in closed form whilst the moment generating function, for example, does. Thus, estimation based on maximum likelihood is difficult. Lastly, as we shall see, standard methods for calculating IFs are woefully inadequate. Without the extensions we have made it would be practically impossible to evaluate them. Indeed, it was this problem which spurred the the need for more sophisticated influence techniques.

In section 2 we examine the distribution theory for the Lagged-Normal model. Section 3 is concerned with estimation and influence theory for three suggested transform methods. We employ the method-of-moments and seek to find explicit parameter estimators. The purpose of section 4 is to make comparisons with maximum likelihood. This necessarily leads us to considerations of efficiency and robustness. We also investigate the performance of estimators in the presence of contamination, via a simulation study. In the final section we briefly consider the influence properties of the ISE technique in this context.

2. Distribution Theory.

Suppose that $X \sim N(\mu,\sigma^2)$ and $Y_i \sim \exp(\lambda_i)$, $i=1,...,r$, then

$$Z = X+\sum_{i=1}^{r} Y_i \tag{4.1}$$

is said to have a *Lagged-Normal* distribution. This definition makes clear the motivation for the name of this distribution, since it is found as a positive number added to a Normal random variable. In this thesis we focus on the case $\lambda_i \equiv \lambda$, so that Z is the sum of Normal and Gamma, denoted Y, random variables.

The density function of Z cannot be written-down explicitly. Denoting it by $f_Z(.)$, the convolution integral is

$$f_Z(z) = \frac{\lambda^r}{(2\pi)^{1/2}\sigma\,\Gamma(r)} \int_0^\infty y^{r-1} e^{-[(z-y-\mu)^2+2\lambda\sigma^2 y]/(2\sigma^2)}\,dy\ . \tag{4.2}$$

For the case $r=1$, it follows readily that

$$f_Z(z) = \lambda e^{\lambda^2\sigma^2/2-\lambda(z-\mu)}\,\Phi[\sigma^{-1}(z-\mu-\lambda\sigma^2)]\ , \tag{4.3}$$

where $\Phi(.)$ denotes the c.d.f. of the $N(0,1)$ distribution.

Despite this lack of a closed-form expression for the density, we can readily find one for the moment generating function (m.g.f.), $M_Z(.)$ say. Assuming X and Y to be independent,

$$M_Z(s) = M_X(s)M_Y(s)$$

in obvious notation. Thus,

$$M_Z(s) = (1-s\lambda^{-1})^{-r} e^{s\mu+s^2\sigma^2/2} \qquad s<\lambda\ . \tag{4.4}$$

In this chapter we seek to exploit this explicit form for the m.g.f. in estimation, concentrating on the robustness of resulting estimators in particular.

3. Influence and Estimation Theory for Transform Methods.

We employ the method-of-moments in this work in an attempt to find explicit parameter estimators. Three methods are investigated, although they form by no means an exhaustive set. In the first we look at direct moment-estimation, carried-out by simply equating empirical and

theoretical transforms at 4 distinct values of s. The second approach uses non-standard transforms, and includes more theoretical refinements for calculating influence functions. The final method is based on characteristic functions, easily found by replacing (s) by (is) in the m.g.f..

3.1. Direct Moment Estimation.

Given a random sample $Z_1, \ldots, Z_n$ from the Lagged-Normal distribution, we form estimating equations as

$$\hat{M}_Z(s_i) = M_Z(s_i) \qquad i=1,2,3,4. \tag{4.5}$$

We find the empirical transform

$$\hat{M}_Z(s) = n^{-1} \sum_{j=1}^{n} e^{sZ_j}, \tag{4.6}$$

on appealing to equation (3.2).

Influence functions for the estimators follow readily. Introducing the notation

$$\begin{aligned} h(s_i) &= h_i = 1 - \lambda^{-1} s_i \\ q(s_i) &= q_i = s_i\mu + s_i^2\sigma^2/2, \end{aligned} \tag{4.7}$$

then to find ∇ we calculate

$$\begin{aligned} \frac{\partial M(s)}{\partial r} &= -h^{-r}\log h s^q \\ \frac{\partial M(s)}{\partial \lambda} &= -rs\lambda^{-2}h^{-(r+1)}e^q \\ \frac{\partial M(s)}{\partial \mu} &= sh^{-r}e^q \\ \frac{\partial M(s)}{\partial \sigma^2} &= 2^{-1}s^2h^{-r}e^q. \end{aligned} \tag{4.8}$$

Thus, the ith row of ∇ is

$$[-h_i^{-r}\log h_i e^{q_i}, -s_i r\lambda^{-2}h_i^{-(r+1)}e^{q_i}, s_i h_i^{-r}e^{q_i}, 2^{-1}s_i^2 h_i^{-r}e^{q_i}] \quad i=1,2,3,4. \tag{4.9}$$

The vector $IF(X; \underline{\hat{M}})$ follows by elementary methods, so our main task is to invert ∇. Since this Jacobian matrix has common row and column factors, we may factorize it as

$$\nabla = D_1 H D_2,$$

where

$$D_1 = diag\ (-h_i^{-r} e^{q_i}) \quad i=1,...,4$$
$$D_2 = diag\ (1, r\lambda^{-2}, 1, 2^{-1})\ .$$

Thus,

$$\nabla^{-1} = D_2^{-1} H^{-1} D_1^{-1}\ . \tag{4.10}$$

This reduces the task to calculating H^{-1}, a somewhat easier task than finding ∇^{-1}. However, the general form is still very complicated and not very informative. Despite this, our theoretical work tells us that individual influence functions may be expressed as

$$IF(X; \hat{\theta}_i) = \sum_{j=1}^{4} k_{ij}(\Theta, \underline{s})[e^{s_j X} - M(s_j)] \quad i=1,...,4, \tag{4.11}$$

where $\Theta = (r, \lambda, \mu, \sigma^2)'$ and the k_{ij} are of complicated form, but do not depend on X. The estimators are not, therefore, influence-bounded robust; this is a consequence of basing estimation on moment generating functions.

Turning now to estimation, we note that $\hat{\lambda}$ is confounded with $\hat{r}$. It is not possible, therefore, to obtain explicit estimators. Because of these problems, we go on to look at an alternative method in the next section which does produce explicit estimators. To achieve this we have to employ non-standard transforms.

3.2. A Derivative-Based Approach.

In the spirit outlined above, we define the modified transform

$$\psi_Z(s) = \frac{\partial}{\partial s} \log M_Z(s)\ , \tag{4.12}$$

so that

$$\psi_Z(s) = \mu + s\sigma^2 + r\ (\lambda - s)^{-1}\ . \tag{4.13}$$

We form estimating equations, as justified in Appendix II, by taking successive derivatives, viz:

$$\hat{\psi}_Z^{(i-1)}(s) = \psi_Z^{(i-1)}(s; \hat{\Theta}) \quad i=1,...,4. \tag{4.14}$$

Explicitly, the estimating equations are

$$\hat{\psi}(s) = \hat{\mu} + s\hat{\sigma^2} + \hat{r}\ (\hat{\lambda} - s)^{-1}$$

$$\hat{\psi}^{(1)}(s) = \hat{\sigma^2} + \hat{r}\ (\hat{\lambda} - s)^{-2}$$

$$\hat{\psi}^{(2)}(s) = 2\hat{r}\ (\hat{\lambda} - s)^{-3}$$

$$\hat{\psi}^{(3)}(s) = 6\hat{r}\ (\hat{\lambda} - s)^{-4}\ ,$$

yielding the estimators

$$\hat{\mu} = \hat{\psi} - s[\hat{\psi}^{(1)} - 1.5\ (\hat{\psi}^{(2)})^2 / \hat{\psi}^{(3)}] - 4.5\ (\hat{\psi}^{(2)})^3 / (\hat{\psi}^{(3)})^2$$

$$\hat{r} = 13.5\ (\hat{\psi}^{(2)})^4 / (\hat{\psi}^{(3)})^3$$

$$\hat{\lambda} = s + 3\ \hat{\psi}^{(2)} / \hat{\psi}^{(3)}$$

$$\hat{\sigma^2} = \hat{\psi}^{(1)} - 1.5\ (\hat{\psi}^{(2)}) / \hat{\psi}^{(3)}\ . \tag{4.15}$$

So in order to estimate the parameters of this model, we must be able to form estimates of the ψs. These depend on the moment generating function, $M_Z(.)$, and its derivatives. We may, therefore, form estimates of the ψs from transform estimators of the usual type.

The first approach adopted to find the influence functions for these estimators was based on the methods of Chapter 2. However, this became too unwieldy and impractical. Indeed, it was impossible to have any degree of confidence in the results so obtained due to the complexity. Also, results were difficult to simplify and interpret. This provided the impetus to develop the methods of Chapter 3.

When employing direct moment estimation, it is sufficient to calculate the first row of ∇, the remaining rows obtained by symmetry. This does not follow here, since each row of the Jacobian matrix corresponds to a different transform. As such, differentiation must be carried-out element-by-element. The task we face is eased by the choice of transform, which gives a number of zero cells in ∇. We find

$$\nabla = \begin{bmatrix} (\lambda - s)^{-1} & -r\ (\lambda - s)^{-2} & 1 & s \\ (\lambda - s)^{-2} & -2r\ (\lambda - s)^{-3} & 0 & 1 \\ 2(\lambda - s)^{-3} & -6r\ (\lambda - s)^{-4} & 0 & 0 \\ 6(\lambda - s)^{-4} & -24r\ (\lambda - s)^{-5} & 0 & 0 \end{bmatrix}. \tag{4.16}$$

To find $IF(X;\hat{\Theta})$, we must first find $IF(X;\hat{\Psi})$, where $\Psi = (\psi, \psi^{(1)},..., \psi^{(3)})'$. This is not an elementary calculation, but the following result enables us to exploit the structure of Ψ to find this quantity.

Theorem 4.1.

Given that the elements of Ψ are formed iteratively by successive differentiation, then we have

$$IF(X;\hat{\psi}^{(k+1)}) = \frac{\partial}{\partial s} IF(X;\hat{\psi}^{(k)}) \quad k=0,1,2.$$

□

Proof

By definition (2.6),

$$IF(X;\hat{\psi}^{(k+1)}) = \frac{\partial}{\partial \varepsilon}\psi^{(k+1)}[(1-\varepsilon)F+\varepsilon A],$$

evaluated at $\varepsilon = 0$. We have the relation that

$$\hat{\psi}^{(k+1)} = \frac{\partial}{\partial s}\hat{\psi}^{(k)} \quad k=0,1,2.$$

Thus,

$$IF(X;\hat{\psi}^{(k+1)}) = \frac{\partial}{\partial \varepsilon}\frac{\partial \psi^{(k)}}{\partial s}[(1-\varepsilon)F+\varepsilon A].$$

Before we can proceed with this proof, we require the following

Lemma

Given the conditions

(i) $\dfrac{\partial \psi^{(k)}}{\partial s}$ and $\dfrac{\partial \psi^{(k)}}{\partial \varepsilon}$ are continuous;

(ii) $\dfrac{\partial^2 \psi^{(k)}}{\partial \varepsilon\, \partial s}$ exists and is continuous, then

$$\frac{\partial^2 \psi^{(k)}}{\partial \varepsilon \, \partial s} = \frac{\partial^2 \psi^{(k)}}{\partial s \partial \varepsilon} \, .$$

For more details, see Delillo (1982) pp 181, Theorem 3.26.

□

The conditions of the Lemma are satisfied here, so we may complete the proof of this theorem. Reversing the order of differentiation in (4.16), we obtain

$$IF(X; \hat{\psi}^{(k+1)}) = \frac{\partial}{\partial s} \frac{\partial}{\partial \varepsilon} \psi^{(k)} [(1-\varepsilon)F + \varepsilon A]$$

□

Using this theorem, we first calculate $IF(X; \hat{\psi})$ and then the others are found iteratively. Now, recalling (4.12), elementary methods yield

$$IF(X; \log \hat{M}(s)) = M(s)^{-1} \, e^{sX} - 1 \, . \tag{4.17}$$

Applying Theorem 4.1,

$$\begin{aligned} IF(X; \hat{\psi}) &= \frac{\partial}{\partial s} [M(s)^{-1} e^{sX} - 1] \\ &= e^{sX} M(s)^{-2} [sM(s) - M'(s)] \, . \end{aligned} \tag{4.18}$$

We now have all the elements in place to investigate the influence behaviour of the estimators given by equations (4.15). Rather than deal in purely general terms, we continue our discussion for the parameter combination $\Theta = (1,1,0,1)'$. We demonstrate the calculation of influence functions and choice of s via generalized variance, employing the symbolic algebra package MAPLE explicitly in this work. As a first step to obtaining $IF(X; \hat{\Theta})$ we obtain the influence vector for the empirical transforms $\hat{\Psi}$. The MAPLE routine (A3.1) of Appendix III is designed to do this, with comments alongside to aid understanding of the MAPLE language. From this we find

$$IF(X; \hat{\psi}^{(i-1)}) = e^{sX - s^2/2} \sum_{j=0}^{i} a_{ji}(s) X^j \tag{4.19}$$

for coefficients $[a_{ji}(s)]$.

As a check on our we working, we now proceed to verify that

$E[IF(X;\hat{\Psi})]=\underline{0}$.

To do this, we need to calculate quantities like $E[X^l e^{sX}]$, l=1,...,5. This is easily done since

$$E(X^l e^{sX}) = \frac{\partial^l}{\partial s^l} M(s), \tag{4.20}$$

which we can find using using MAPLE, the code for which is given in (A3.2).

A difficulty of this method now arises. We have to substitute the expectations for the random variables in $IF(X;\Psi)$, and this cannot be done directly. The reason for this is that MAPLE would not allow direct substitution for quantities such as X^j. The polynomials had to be recoded so that $X^j \rightarrow Xj$. In this way we can substitute the expected quantities. This involved a considerable amount of work, but did lead to the desired verification.

We would naturally go on now to look at $IF(X;\hat{\Theta})$, but this depends on s. As such, we first discuss the choice of the transform variable by minimizing generalized variance. This leads us naturally to consider matters of asymptotic variance. Having done this, we will then be in a position to return to considerations of influence.

In calculating $var(\hat{\Psi})$, the dispersion matrix of the empirical transforms, we still have to recode powers of X. This is a far more time-consuming task than before. If we consider the previous problem to be of order 4, then the present one is of order $4^2 = 16$. This problem would have to be resolved for this technique to be of general use. Despite this, MAPLE has proved to be a valuable tool.

The first step is to calculate the matrix of products and cross-products of influence functions. Recalling the form of (4.19), this matrix will have a common factor e^{2sX-s^2}. This simplifies the working, since we need only consider the polynomial components of the influence functions. The routine (A3.3) generates the matrix *prod* containing these elements. We then have the unpleasant task of re-coding the 10 distinct elements, taking advantage of the symmetric nature of *prod* which saves us 6 recodings.

The expectations we need to find in this context are of the form

$$E[X^l e^{2sX}] = 2^{-l}\frac{\partial^l}{\partial s^l} M(2s) \quad l=0,...,8. \tag{4.21}$$

Denoting this vector of expectations by *eyl*2, we employ MAPLE routine (A3.4) to find its elements.

To be able to achieve the objectives of this section we have left only to find the Jacobian matrix ∇. Substituting for Θ in (4.16) and writing $t = (1-s)^{-1}$, we obtain

$$\nabla = \begin{bmatrix} t & -t^2 & 1 & s \\ t^2 & -2t^3 & 0 & 1 \\ 2t^3 & -6t^4 & 0 & 0 \\ 6t^4 & -24t^5 & 0 & 0 \end{bmatrix} . \tag{4.22}$$

It follows readily using the linear algebra functions of MAPLE that

$$|\nabla| = -18(1-s)^{-8} . \tag{4.23}$$

Recalling result (3.23), the generalized variance function for this example is

$$gv(s) \propto (1-s)^{16} \, |var(\hat{\Psi}(s))| . \tag{4.24}$$

Before we seek to optimize $gv(.)$, note that there are restrictions on the range of s. Since $M(u)$ is valid only for $u < 1$, and we require $M(s)$ as well as $M(2s)$ to be well-defined, we consider only $s < 0.5$. As a first step to locating s^* in this region, we scan the interval $[-1.00, 0.45]$, which may be done via routine (A3.5). The results so obtained suggested that we concentrate on $s \in (-0.20, -0.10)$. Using essentially the same routine as above, values of $gv(.)$ were obtained across this interval to a finer resolution than before. These were then scaled to lie between 0 and 1, yielding Figure 4.1 below. It seems from this that we should choose $s^* \approx -0.166$.

We are now in a position to evaluate the influence functions explicitly. These follow by applying Theorem 3.2, yielding

$$IF(X; \hat{r}) = e^{-0.166X}[-1.063X^4 + 6.59X^3 + 0.452X^2 - 22.3X + 2.51] , \tag{4.25}$$

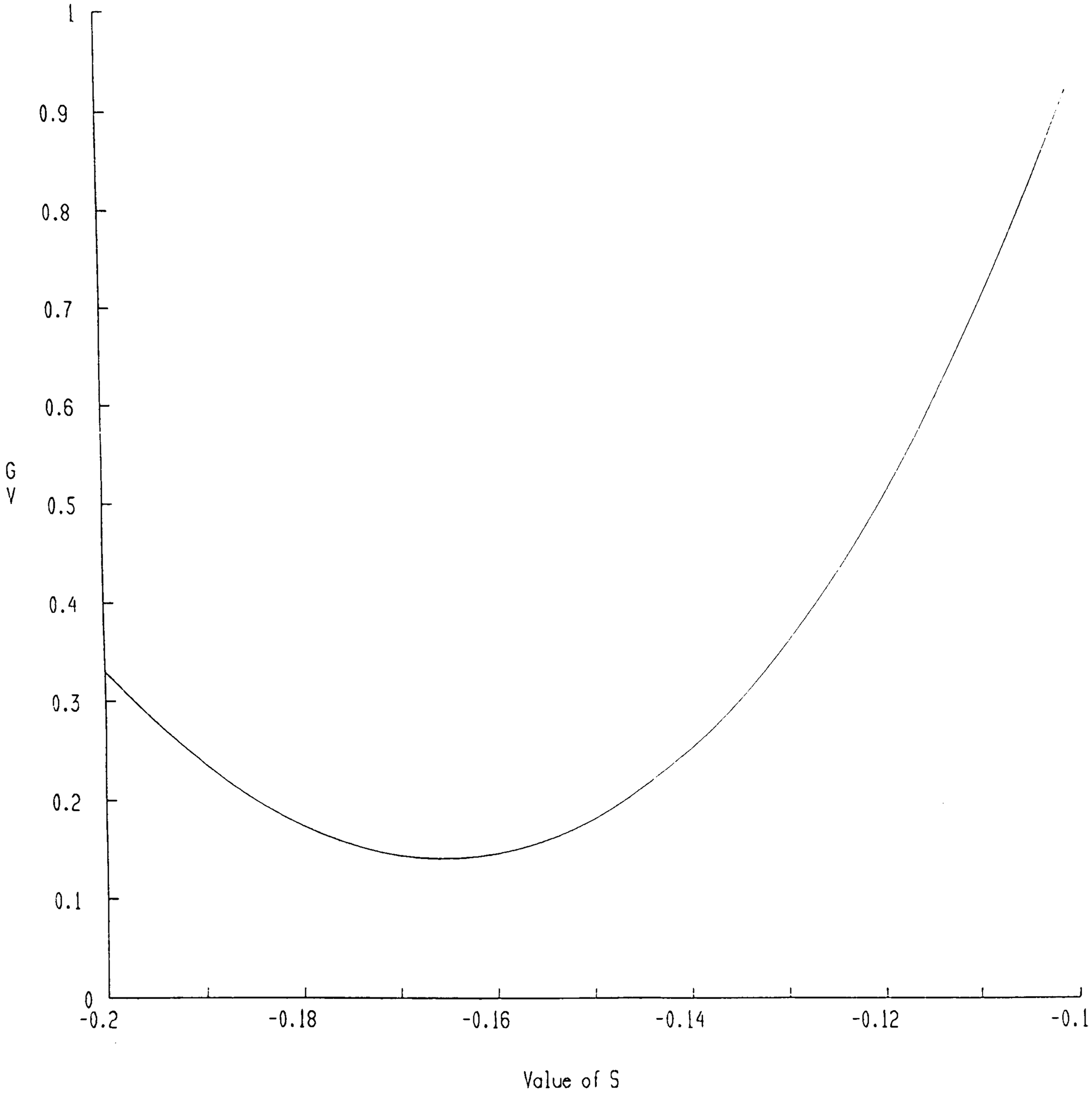
$$IF(X; \hat{\lambda}) = e^{-0.166X}[-0.413X^4 + 2.21X^3 + 0.911X^2 - 7.33X + 0.264] , \tag{4.26}$$

$$IF(X; \hat{\mu}) = e^{-0.166X}[0.651X^4 - 4.37X^3 + 0.609X^2 + 15.8X - 3.25] , \tag{4.27}$$

$$IF(X; \hat{\sigma^2}) = e^{-0.166X}[0.261X^4 - 2.06X^3 + 1.97X^2 + 5.56X - 2.96] , \tag{4.28}$$

with coefficients recorded to 3sf.

Figure 4.1. Scaled Generalized Variance.



These influence functions are depicted graphically in Figures 4.2-4.5 below. We note first that none of these estimators are IBR. However, their behaviour diverges for positive and negative values of X :

$$IF(X;\hat{\theta}_i) = e^{-0.166X} P_i(X) \quad i=1,..,4, \tag{4.29}$$

where $P_i(X)$ denotes a polynomial of degree 4. Thus,

$$\lim_{X\to\infty} |IF(X;\hat{\theta}_i)| = 0,$$
$$\lim_{X\to-\infty} |IF(X;\hat{\theta}_i)| = \infty. \tag{4.30}$$

We can understand this by returning to first principles. We know that

$$IF(X;\hat{M}(s)) = e^{sX} - M(s).$$

If X can take values on the whole of R, then the influence behaviour of $\hat{M}(s)$ depends critically on the sign of s. We observed in equations (4.30) what happens when $s<0$, and the situation is reversed when $s>0$.

Interestingly, for the case we are investigating most probability is allocated to positive values of X. Perhaps, then, it is unsurprising that a negative choice for s^* should emerge.

Turning now to the influence functions individually, the segments depicted are those where greatest variation was observed. Outside these regions, influence behaviour can be inferred from the above discussion in conjunction with the explicit forms (4.28). For $\hat{\mu}$, we note turning points within the usual range of values associated with this Lagged-Normal distribution. Large negative values of X have high influence as we would expect, but positive values can as well, because the damping due to the leading negative-exponential term takes effect only slowly. $\hat{\sigma^2}$ seems more robust than $\hat{\mu}$, an estimator that will receive further attention later on. Large outliers tend to increase the value of $\hat{\sigma^2}$, as we would expect. The behaviour for $\hat{\mu}$ as $X\to-\infty$ makes rather less sense, since we would expect $\hat{\mu}$ to fall rather than rise. However, being of an essentially polynomial nature, it must behave the same for both extremes. This is an unappealing feature of estimation based on m.g.f.s, due to their basic lack of symmetry in X.

Turning now to the parameters of the gamma distribution, we witness a very different pattern of behaviour for $\hat{r}$ and $\hat{\lambda}$. Very large outliers tend to decrease both of these, although rather

Figure 4.2. Influence Function for $\hat{\mu}$.

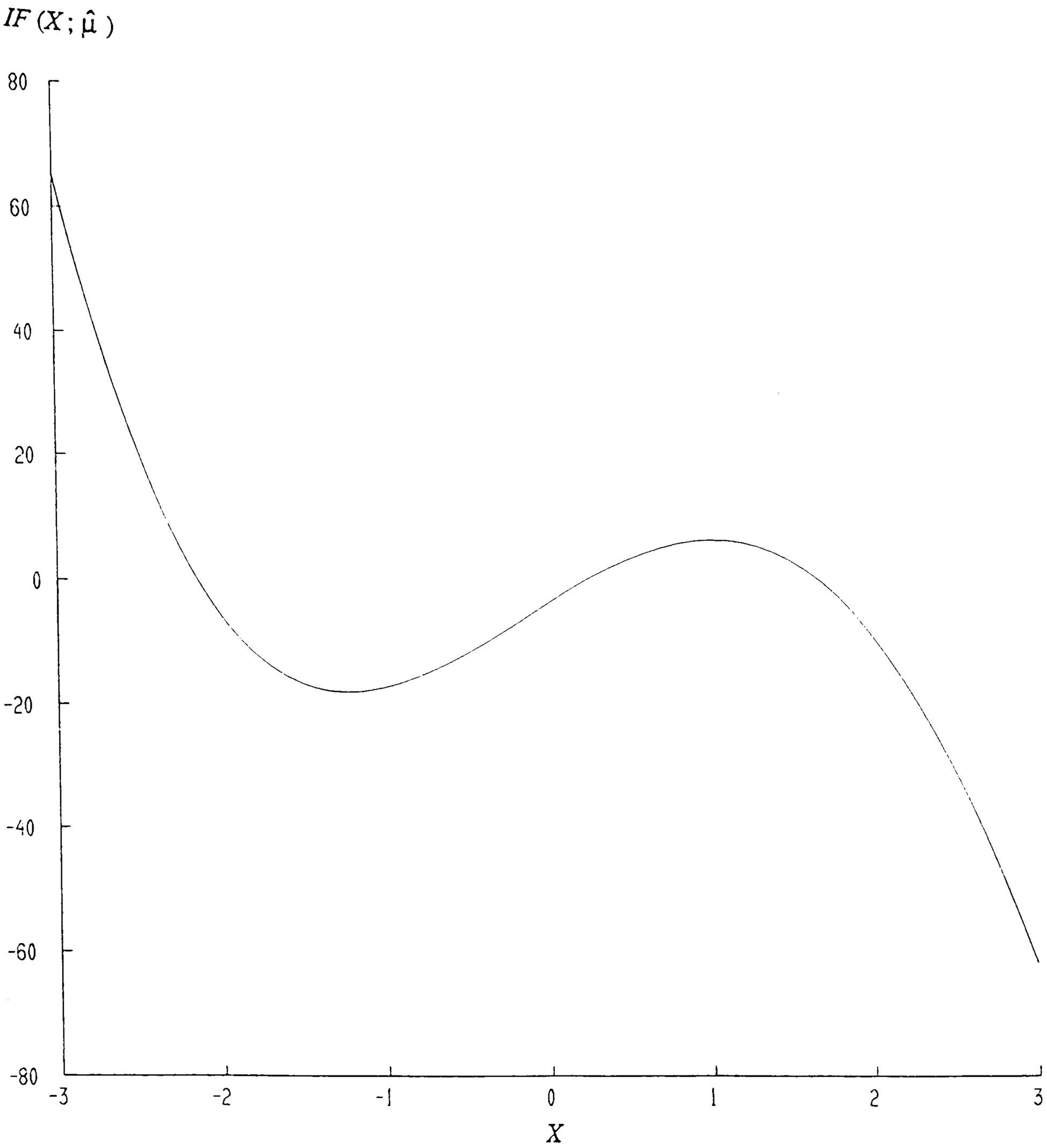


Figure 4.3. Influence Function for $\hat{\sigma^2}$.

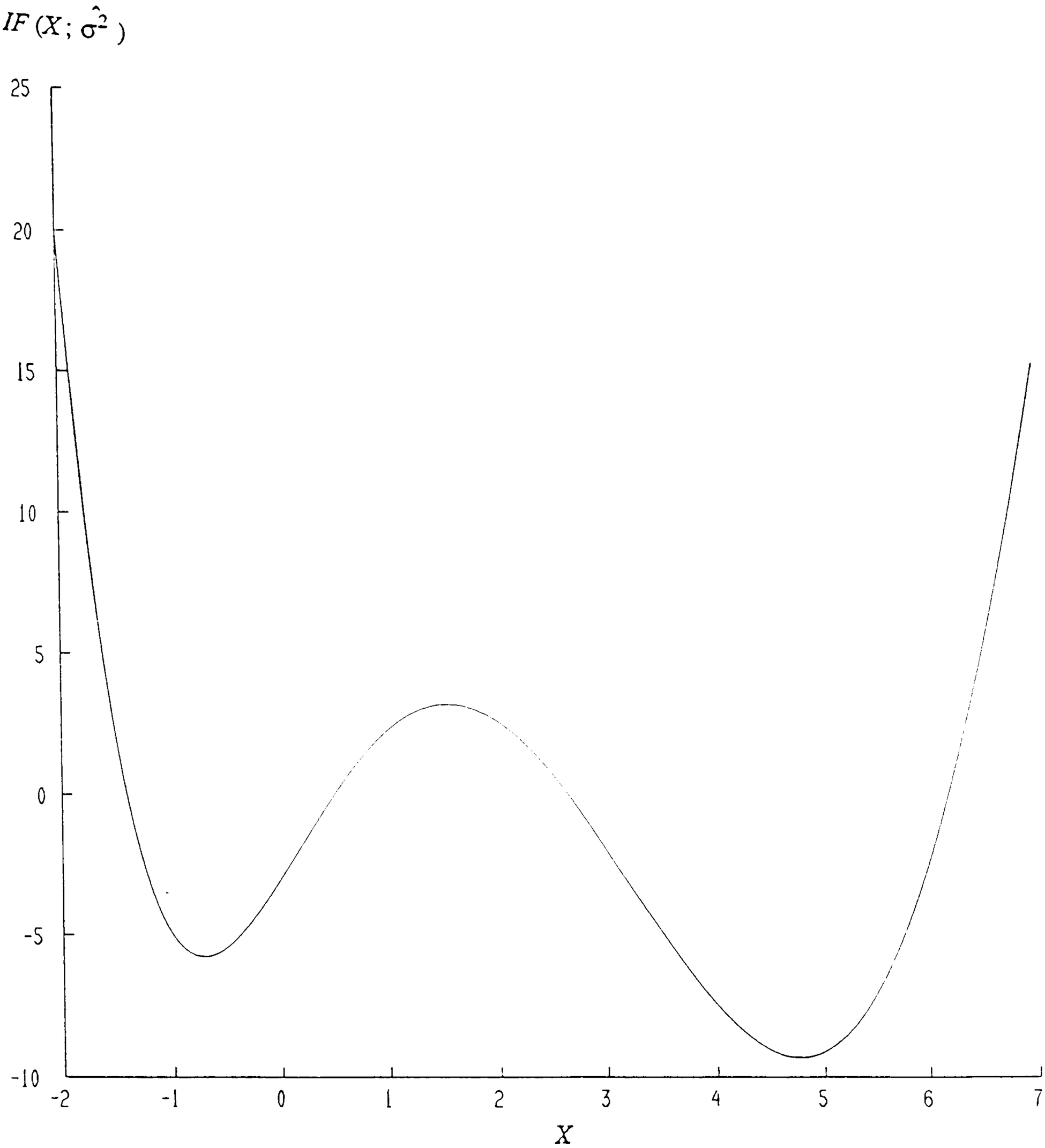


Figure 4.4. Influence Function for $\hat{r}$.

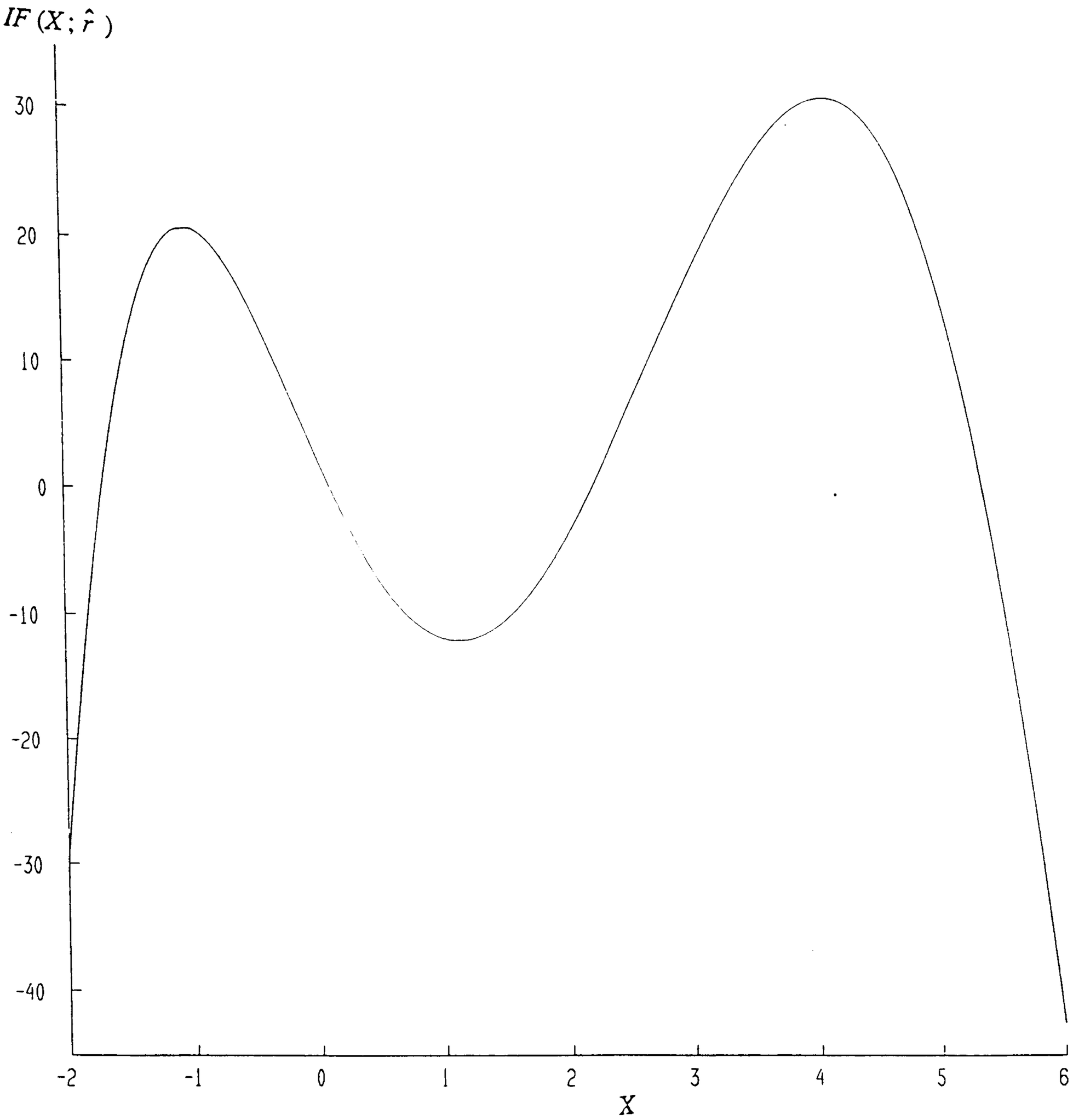
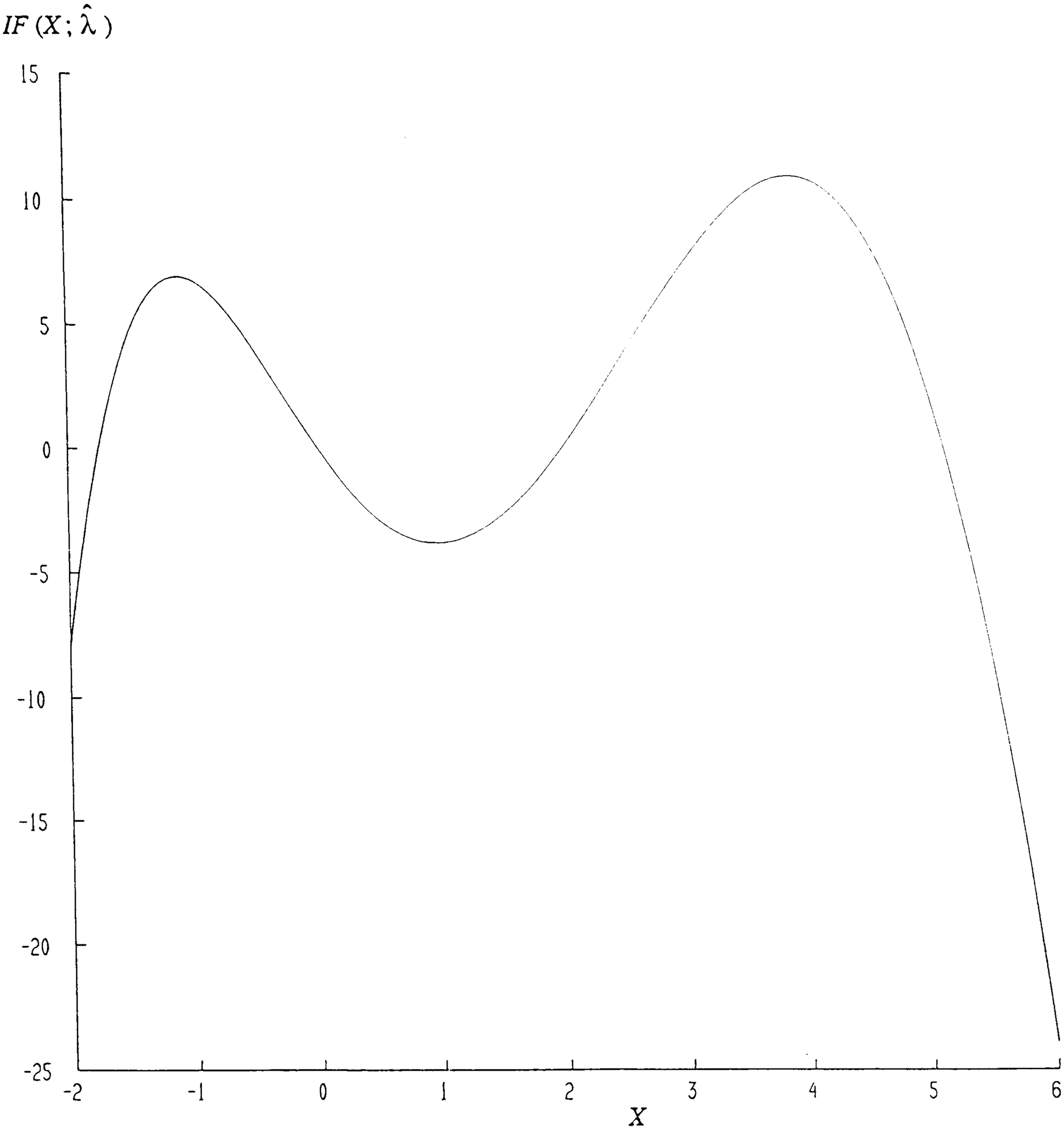


Figure 4.5. Influence Function for $\hat{\lambda}$.



more sharply in the case of $\hat{r}$. Again, this is somewhat counter-intuitive. There are two reasons for this. First, we have to take account of the asymmetry discussed earlier. Second, both $\hat{r}$ and $\hat{\lambda}$ are very strongly negatively-correlated with $\hat{\mu}$, having correlations of -0.99 and -0.92 respectively. These statistics were calculated from considerations of asymptotic variance, which we discuss next.

First, we find the asymptotic result

$$var(\hat{\Psi}) \approx n^{-1} \begin{bmatrix} 1.719 & & & \\ 0.02046 & 6.660 & & \\ 0.8778 & 7.836 & 62.31 & \\ 1.617 & 30.55 & 217.8 & 1446 \end{bmatrix}. \tag{4.31}$$

(Since this matrix is symmetric we only quote the lower triangle, correct to 4sf.) We note that the empirical transforms grow progressively poorer as the order of differentiation increases. This is a phenomenon similar to that of conventional moment estimation, where the use of higher-order moments tends to lead to unstable estimators. We note the particularly poor performance of $\hat{\psi}^{(3)}$.

Unsurprisingly, as we can see below, this carries-over to parameter estimation, particularly for r and μ.

$$var(\hat{\Theta}) \approx n^{-1} \begin{bmatrix} 585.3 & & & \\ 228.6 & 95.22 & & \\ -356.7 & -134.4 & 223.3 & \\ -145.5 & -51.77 & 93.69 & 46.11 \end{bmatrix}. \tag{4.32}$$

We note the strong correlations quoted above. The results here suggest that this method of estimation performs very poorly in terms of asymptotic variance, and we go on now in the next section to discuss the underlying reasons.

3.2.1. Explanation of Poor Performance.

The estimators of r and μ perform particularly badly, where powers of $\hat{\psi}^{(3)}$ appear in the denominator, as can be seen by referring to equations (4.15). In a small simulation study, some values of $\hat{\mu}$ were recorded which showed that ridiculous estimates could be produced. For example, values of the order -4420 were not uncommon. The reason for this seems to be that $\hat{\psi}^{(3)}$ can

often take values close to zero. Table 4.1 below summarizes the results of a simulation study of $\hat{\psi}^{(3)}$. We see that if we were to form approximate 95% confidence intervals for $\psi^{(3)}$, then 0 would lie comfortably within these intervals for n up to 100. We conclude from this that $\hat{\psi}^{(3)}$ has a strongly destabilizing effect on this estimation procedure, echoing the similarity with conventional moment estimation noted above.

Table 4.1. Simulation Study of $\hat{\psi}^{(3)}$.

Sample Statistics:		
n	Mean	Variance
20	1.284	78.041
100	0.782	17.875
500	0.711	3.642
1000	0.713	1.806

Rounding to 3dp.

We note in passing that the sample variance seems to be converging to its theoretical limit (4.31) very slowly.

The question which arises now is whether, assuming one parameter to be known, we can significantly improve this method by avoiding the third derivative. It seems reasonable to assume that in many cases r would be known, or pre-selected in some fashion. If this is the case, then we obtain the estimators

$$\hat{\mu} = \hat{\psi} - r[2r/\hat{\psi}^{(2)}]^{-1/3} - s^*[\hat{\psi}^{(1)} - r\{2r/\hat{\psi}^{(2)}\}^{-2/3}]$$

$$\hat{\sigma^2} = \hat{\psi}^{(1)} - r[2r/\hat{\psi}^{(2)}]^{-2/3}$$

$$\hat{\lambda} = s^* + [2r/\hat{\psi}^{(2)}]^{1/3} . \tag{4.33}$$

The table below summarizes the results of a simulation study of $\hat{\psi}^{(2)}$, showing that the problems with near-zero values are less severe, but that it is still not a very stable transform on which to

base estimation procedures.

Table 4.2. Simulation Study of $\hat{\psi}^{(2)}$.

Sample Statistics:		
n	Mean	Variance
20	1.304	3.028
100	1.300	0.716
500	1.313	0.138
1000	1.312	0.069

Rounding to 3dp.

The results of this section show that this method does not provide a practical estimation technique for the Lagged-Normal distribution. However, it is a useful demonstration of influence techniques allied in practice with symbolic algebra.

3.3. Method-of-Moments Based on Characteristic Functions.

In this section we again assume that r is known, leaving a vector of unknown parameters $\Theta = (\lambda,\mu,\sigma^2)'$ to estimate. The appropriate characteristic function is

$$\phi_Z(s) = M_Z(is).$$

Recalling (4.4), we may write this as the complex exponential

$$\phi_Z(s) = (1+s^2\lambda^2)^{-r/2} e^{-s^2\sigma^2/2}\, e^{i(s\mu + r\tan^{-1}(s/\lambda))}. \tag{4.34}$$

Taking two values of s, s_1 and s_2, we may form the estimating equations

$$\hat{R}_j = R_j \quad j=1,2;$$
$$\hat{I}_j = I_j \quad j=1,2.$$

Since we only have three parameters to estimate, we combine the equations in s_2 to give three estimating equations in three unknowns, viz:

$$\begin{aligned} \hat{R}_1 &= R_1 \\ \hat{I}_1 &= I_1 \\ \hat{I}_2\,\hat{R}_2^{-1} &= I_2\,R_2^{-1}\ , \end{aligned} \tag{4.35}$$

in notation established in Chapter 3. We solve for $\hat{\lambda}$ and $\hat{\mu}$ in the system

$$\begin{aligned} \hat{I}_1\,\hat{R}_1^{-1} &= I_1\,R_1^{-1} \\ \hat{I}_2\,\hat{R}_2^{-1} &= I_2\,R_2^{-1}\ . \end{aligned} \tag{4.36}$$

Writing

$$\hat{\kappa}_i = \tan^{-1}(\hat{I}_i\,\hat{R}_i^{-1}) \quad i=1,2, \tag{4.37}$$

we obtain

$$s_2\hat{\kappa}_1 - s_1\hat{\kappa}_2 = r[s_2\tan^{-1}(s_1\hat{\lambda}^{-1}) - s_1\tan^{-1}(s_2\hat{\lambda}^{-1})]\ . \tag{4.38}$$

Once $\underline{s}^*$ has been chosen from some external criterion, such as minimizing generalized variance, we may solve (4.38) for $\hat{\lambda}$ by simple numerical techniques. Given this has been done, we obtain $\hat{\mu}$ and $\hat{\sigma}^2$ by back-substitution.

Even though this procedure does not produce explicit estimators, we can select $\underline{s}$ and investigate the robustness properties of these estimators without great difficulty. Looking first at influence, the only new empirical transform is $\hat{\psi} = \hat{I}_2\,\hat{R}_2^{-1}$, the influence function for which is easily derived as

$$\begin{aligned} IF(X;\hat{\psi}) &= R_2^{-1}IF(X;\hat{I}_2) - I_2R_2^{-2}IF(X;\hat{R}_2) \\ &= R_2^{-1}\sin(s_2X) - I_2R_2^{-1}\cos(s_2X)\ . \end{aligned} \tag{4.39}$$

Given the influence function for $\hat{\psi}$, we can now calculate its asymptotic variance using Result 2.2, as follows.

$$\begin{aligned} var(\hat{\psi}) &\approx n^{-1}\int_{-\infty}^{\infty}[IF(X;\hat{\psi})]^2 dF(X) \\ &= R_2^{-2}var(\hat{I}_2) + I_2^2R_2^{-4}var(\hat{R}_2) - 2I_2R_2^{-3}cov(\hat{I}_2,\hat{R}_2)\ . \end{aligned} \tag{4.40}$$

We may also calculate the asymptotic covariances involving $\hat{\psi}$, viz:

$$\begin{aligned} cov(\hat{\psi},\hat{I}_1) &\approx n^{-1}\int IF(X;\hat{\psi})IF(X;\hat{I}_1)dF(X) \\ &= R_2^{-1}cov(\hat{I}_2,\hat{I}_1) - I_2R_2^{-2}cov(\hat{R}_2,\hat{I}_1)\ . \end{aligned} \tag{4.41}$$

By symmetry we find

$$cov(\hat{\psi},\hat{R}_1) \approx R_2^{-1} cov(\hat{I}_2,\hat{R}_1) - I_2 R_2^{-1} cov(\hat{R}_2,\hat{R}_1)\,. \tag{4.42}$$

All of the quantities to be found now involve only standard transforms, and are therefore straightforward to calculate, so we omit the details here. For the rest of this discussion we set $\mu = 0$ and $r = \lambda = \sigma^2 = 1$. We may proceed along the lines of the previous section to examine the generalized variance function. A rough search yielded $\underline{s}^* \approx (0.3,0.6)'$. For this choice of $\underline{s}$, we obtain the asymptotic dispersion matrix (elements quoted to 4sf):

$$var(\hat{R}_1,\hat{I}_1,\hat{\psi})' \approx n^{-1} \begin{bmatrix} 0.03785 & & \\ -0.04652 & 0.1237 & \\ -0.1242 & 0.3805 & 1.253 \end{bmatrix}. \tag{4.43}$$

This method is clearly a drastic improvement. Note, however, that the modified transform is much less efficient than the transforms of standard form. This appears to be a phenomenon of general application. Employing non-standard, modified transforms is often convenient, particularly in providing explicit estimators. However, these are attained at the cost of a loss of efficiency, which may well be very large.

Applying standard theory, and using MAPLE to calculate the Jacobian matrix, we obtain

$$var(\hat{\lambda},\hat{\mu},\hat{\sigma^2})' \approx n^{-1} \begin{bmatrix} 5.883 & & \\ 4.853 & 5.837 & \\ 4.986 & 4.919 & 9.873 \end{bmatrix}, \tag{4.44}$$

from (4.43).

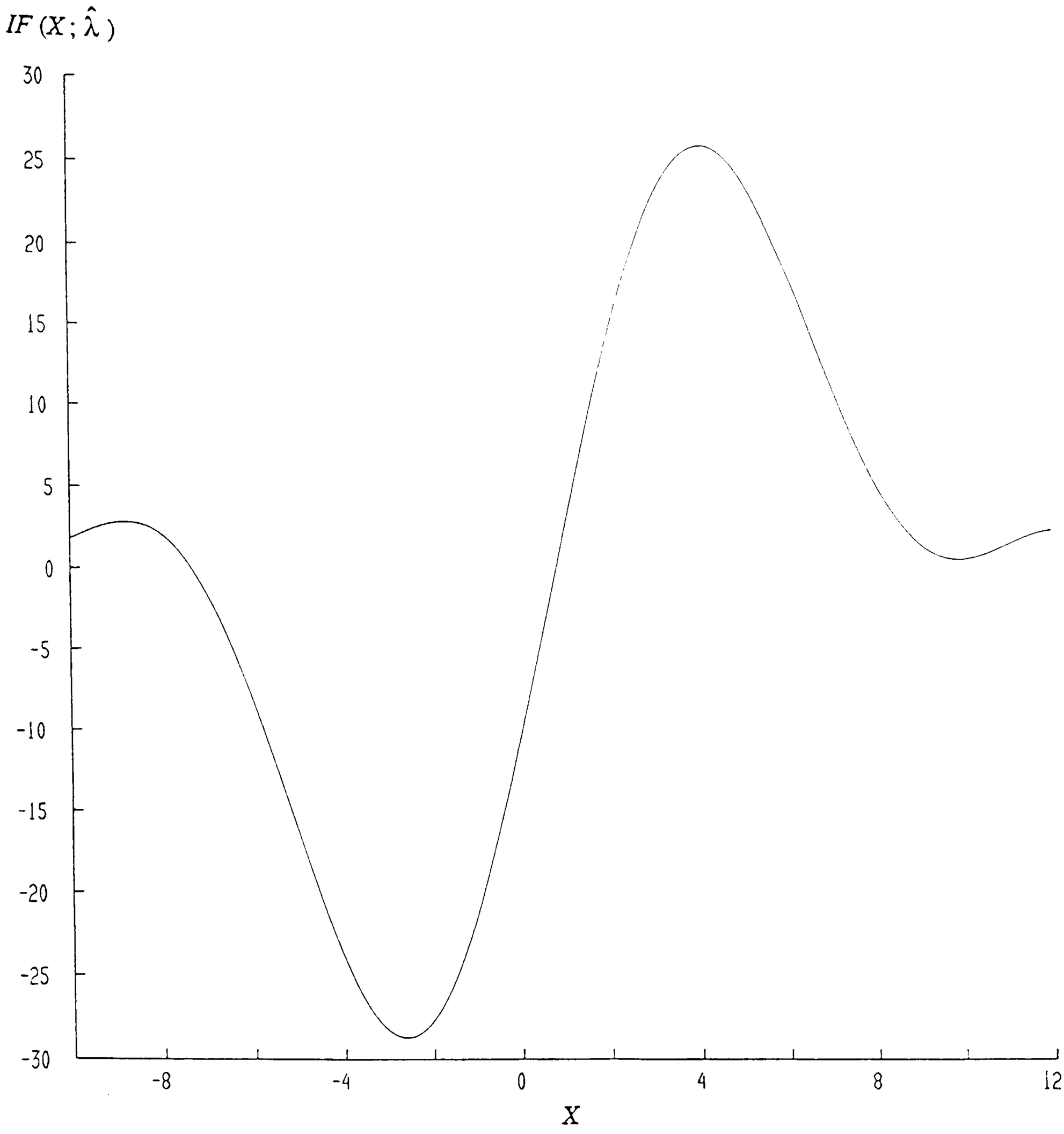
A consequence of basing estimation on the Sine and Cosine transforms of the density is that the resulting estimators are IBR. Further, their influence functions will be linear combinations of sinusoidal terms. A typical case is that for $\hat{\lambda}$:

$$IF(X;\hat{\lambda}) = 20 \sin 0.3(X-1) + 12 \sin 0.2(3X-2)\,, \tag{4.45}$$

where the coefficients have been rounded to the nearest digit. This is depicted graphically below in Figure 4.6.

In section 3 we have investigated the influence properties and statistical performance of some techniques based on the method-of-moments. It seems that it is not possible to obtain explicit estimators which are also efficient due to the awkward form of M_Z (.). This difficulty arises from the very different algebraic forms of the Normal and Exponential moment generating

Figure 4.6. Influence Function of $\hat{\lambda}$.



functions. It may well be worthwhile for future work to focus on applying integrated-squared-error methods to this problem. The integral involved cannot, it seems, be evaluated explicitly. However, this can be overcome by employing an approach based on numerical quadrature, as discussed by Cummings (1983). Our work on ISE in this thesis suggests that this may well be profitable. We move on now to consider the application of maximum likelihood in this setting.

4. Comparisons With Maximum Likelihood.

We re-quote here the density function (4.2) for ease of exposition

$$f_Z(z) = \frac{\lambda^r}{(2\pi)^{1/2}\sigma\,\Gamma(r)} \int_0^\infty y^{r-1} e^{-[(z-y-\mu)^2 + 2\lambda\sigma^2 y]/(2\sigma^2)} dy \ .$$

Given a random sample of observations $Z_1, \ldots, Z_n$ from the Lagged-Normal distribution, we may form the likelihood function in terms of $\Theta = (r, \lambda, \mu, \sigma^2)'$ as

$$l(\Theta) = \lambda^{nr} \{ (2\pi)^{1/2}\sigma\Gamma(r) \}^{-n} \prod_{i=1}^{n} \int_0^\infty y^{r-1} e^{-[(Z_i - y - \mu)^2 + 2\lambda\sigma^2 y]/(2\sigma^2)} \, dy \ . \tag{4.46}$$

Writing

$$k(\Theta) = nr\log\lambda - n/2\,\log\sigma^2 - n\log\Gamma(r)$$

and

$$Q(Z_i; \Theta) = \int_0^\infty y^{r-1} e^{-[(Z_i - y - \mu)^2 + 2\lambda\sigma^2 y]/(2\sigma^2)} \, dy \ ,$$

we may write the log-likelihood as

$$L(\Theta) = k(\Theta) + \sum_{i=1}^{n} \log Q(Z_i; \Theta) \ . \tag{4.47}$$

Given this notation, we are now in a position to move on to the first major subject of this section.

4.1. Efficiency.

The quantities we require here are the second derivatives of the log-likelihood, and then subsequently their expectations. The best lower-bounds for the variances of the Lagged-Normal estimators are provided by the diagonal elements of the inverse information matrix. We may find lower-bounds for these by simply inverting the diagonal elements of the information matrix itself.

The bound is exact when the information elements are uncorrelated. See Cox & Hinkley (1986) pp 255-56 for details. This is the scheme we employ in this work, to get a feel for the efficiency of the characteristic function-based method versus maximum likelihood. We focus attention on efficiency calculations for λ and μ here.

Now, the second derivatives are all of the form

$$\frac{\partial^2 L}{\partial \theta_j^2} = \frac{\partial^2 k}{\partial \theta_j^2} + \sum_{i=1}^{n} \{ Q(Z_i) Q_{\theta_j}^{(2)}(Z_i) - [Q_{\theta_j}^{(1)}(Z_i)]^2 \} Q(Z_i)^{-2} , \tag{4.48}$$

with the θ_j subscript denoting the parameter with respect to which differentiation is carried-out. We may write the above as

$$\frac{\partial^2 L}{\partial \theta_j^2} = \sum_{i=1}^{n} [g_1(Z_i) - g_2(Z_i)] , \tag{4.49}$$

where

$$g_1(Z) = Q_{\theta_j}^{(2)}(Z) / Q(Z)$$

and

$$g_2(Z) = [Q_{\theta_j}^{(1)}(Z) / Q(Z)]^2 .$$

First, the calculations for the Cramér-Rao lower bound for μ. We have to find the expected values of $g_1(Z)$ and $g_2(Z)$, and we may employ the delta method to find approximations to these quantities. Explicitly, we obtain

$$E[g_j(Z_i)] \approx g_j[E(Z_i)] + 1/2\, var(Z_i) \frac{\partial^2 g_j}{\partial Z_i^2} \qquad j=1,2, \tag{4.50}$$

where the partial derivative is evaluated at $E(Z_i)$. In the case of $\Theta = (1,1,0,1)'$, we have

$$E(Z_i) = \mu + r\lambda^{-1} = 0 + 1 = 1$$

and

$$var(Z_i) = \sigma^2 + r\lambda^{-2} = 1 + 1 = 2 ,$$

so that

$$n^{-1} E\left(\frac{\partial^2 L}{\partial \mu^2} \right) = g_1(1) - g_2(1) + \frac{\partial^2 g_1(Z=1)}{\partial Z^2} - \frac{\partial^2 g_2(Z=1)}{\partial Z^2} . \tag{4.51}$$

MAPLE was used extensively to calculate the quantities above, yielding

$$var(\hat{\mu}) \geq 2.272\, n^{-1} \quad (4sf)\,. \tag{4.52}$$

We are therefore able to calculate a lower bound for the asymptotic efficiency of the characteristic function estimator of μ, on referring to (4.44), as 39%. Similar calculations to those above give the efficiency for λ to be greater than 33%.

Clearly, the characteristic function technique performs badly in these terms. However, this is not the whole story. In the next two sections we investigate the robustness of maximum likelihood, where the transform method compares much more favourably.

4.2. Influence.

From the previous section, we have the estimating equations

$$\sum_{i=1}^{n}[Q_{\theta_j}^{(1)}(Z_i)/Q(Z_i) - \frac{\partial k'}{\partial \theta_j}] = 0 \quad j=1,..,4, \tag{(4.53)}$$

where $k' = n^{-1}k$. Appealing to the theory of M-estimation, we obtain

$$IF(X;\hat{\theta}_j) \propto Q_{\theta_j}^{(1)}(X)/Q(X) - \frac{\partial k'}{\partial \theta_j}\,. \tag{4.54}$$

It is not obvious from (4.54) whether the m.l.e.s are IBR or not, so we must investigate the terms involving X a little further. Focusing on the case $\Theta = (1,1,0,1)'$, we obtain from (4.3) that

$$Q(X) = (2\pi)^{1/2} e^{(1-2X)/2}\Phi(X-1)\,. \tag{4.55}$$

For the $Q^{(1)}(.)$ terms,

$$Q_{\lambda}^{(1)}(X) = -\int_0^{\infty} e^{-[(X-y)^2+2y]/2}dy$$

$$Q_{\mu}^{(1)}(X) = \int_0^{\infty}(X-y)e^{-[(X-y)^2+2y]/2}dy$$

$$Q_{\sigma^2}^{(1)}(X) = 1/2\int_0^{\infty}(X-y)^2 e^{-[(X-y)^2+2y]/2}dy\,. \tag{4.56}$$

Employing MAPLE to calculate these, we find that the m.l.e.s $\hat{\mu}_{ml}$ and $\hat{\sigma^2}_{ml}$ are both IBR, but the m.l.e. for λ is not. This is an unappealing feature of maximum likelihood in general, in that it is not directly obvious, for any particular distribution, whether a given m.l.e. is IBR. One has to inspect the estimating equations before a conclusion may be reached; this is not the case

with transforms, as we have seen.

Concentrating on $\hat{\lambda}_{ml}$ for the moment, we find

$$Q_{\lambda}^{(1)}(X) = 1 - e^{-X^2/2} - (2\pi)^{1/2}(X-1)e^{(1-2X)/2}[1-\Phi(X-1)]\,, \tag{4.57}$$

so that

$$IF(X;\hat{\lambda}_{ml}) \propto \frac{1 - e^{-X^2/2} - (2\pi)^{1/2}(X-1)e^{(1-2X)/2}[1-\Phi(X-1)]}{(2\pi)^{1/2}\Phi(X-1)} - 1\,. \tag{4.58}$$

This is clearly bounded as $X \to \infty$, tending, in fact, to $(2\pi)^{-1/2} - 1 \simeq -0.6$. As $X \to -\infty$, however, the influence function is unbounded. Depicted below is that part of the IF which depends on X.

We witness clearly asymmetric behaviour, although in this case it does make some sense. For small values of X, $\hat{\lambda}$ tends to increase, and so reducing, given that other parameters remain fixed, the estimate of the population mean $r\hat{\lambda}^{-1} + \hat{\mu}$. The converse is also true. However, maximum likelihood is clearly sensitive to deviations from the assumed Lagged-Normal model, particularly to outliers below the mean. In the next section we get some idea of what this might mean in practice.

4.3. Behaviour in the Presence of Contamination.

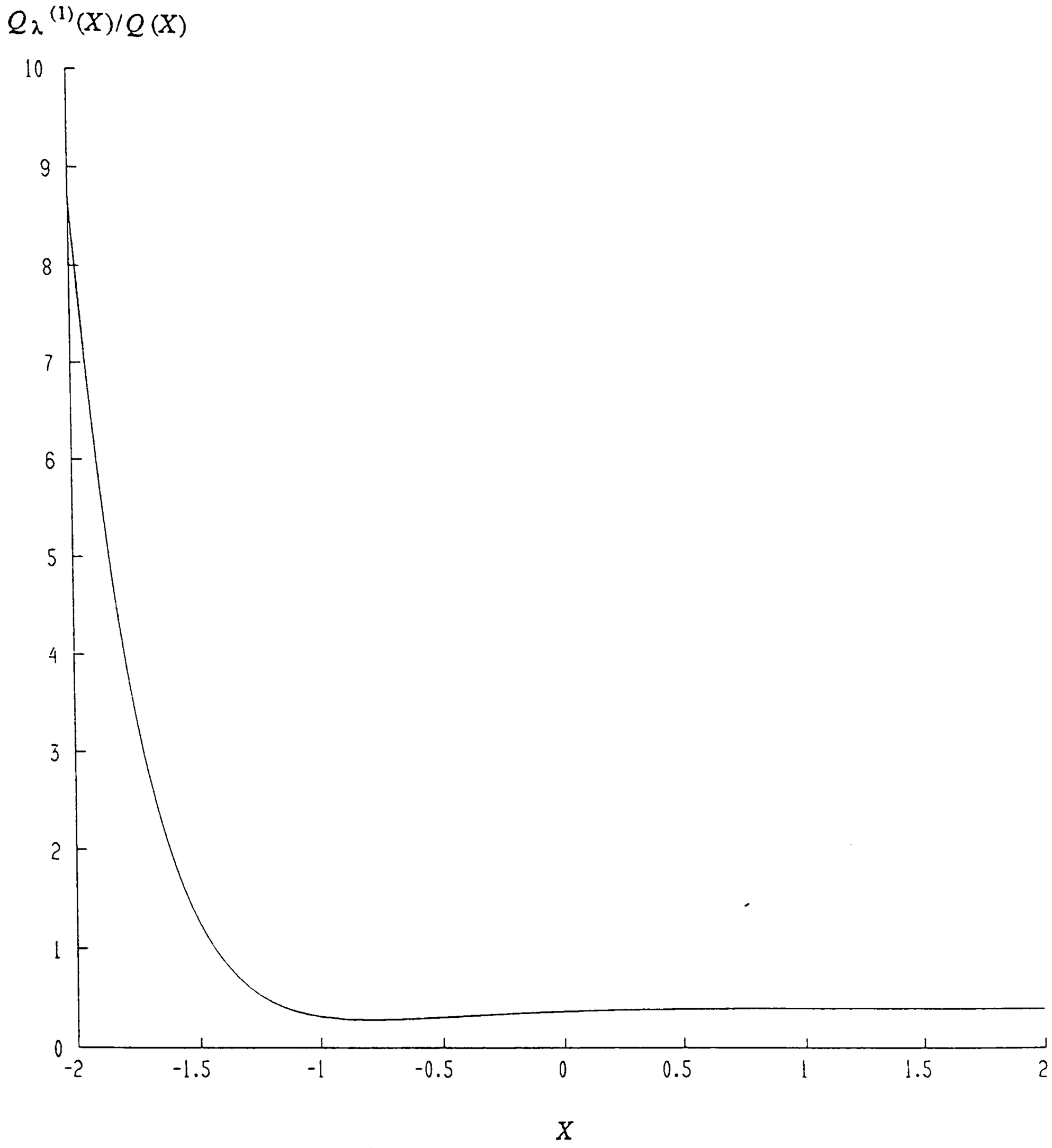
We continue to concentrate on the estimation of λ, but instead of observations following the Lagged-Normal distribution having parameters $\Theta = (1,1,0,1)'$, we in fact observe the contamination model

$$F = (1-\varepsilon)LN(1,1,0,1) + \varepsilon N(0,3^2)\,, \tag{4.59}$$

in obvious notation, denoting the respective c.d.f.s. The objective of this section is to compare the performance of $\hat{\lambda}_{ml}$ versus a characteristic function estimator, $\hat{\lambda}_{cf}$ say.

The NAG routine E04ABF was employed to optimize the likelihood function. For this numerical work it was more convenient to express the standard normal c.d.f. as an error function. Interestingly, for any $\varepsilon > 0$, maximum likelihood experienced severe numerical difficulties. The problem was tracked down to the calculation of $1 + erf[2^{-1/2}(Z_j - \lambda^{(t)})]$, for the t-th iterate. For $Z_j - \lambda^{(t)}$ sufficiently less than zero, this quantity was returned as zero. This lead to a subsequent

Figure 4.7. Plot of $Q_{\lambda}^{(1)}(X)/Q(X)$.



attempt to calculate $log\,(0)$. We know from earlier influence work that $\hat{\lambda}_{ml}$ is sensitive to such observations. This numerical problem was fixed by, somewhat arbitrarily, calculating $log\,(0.01)$ instead. In doing this, we observe an upper-bound on the performance of the m.l.e., which should be borne in mind when interpreting the results to follow. Note that this problem does not arise when $\varepsilon = 0$.

Turning now to $\hat{\lambda}_{cf}$, we base this on the ratio of real and imaginary parts, yielding

$$\hat{\lambda}_{cf}(s) = s\hat{R}\,\hat{I}^{-1}\,. \tag{4.60}$$

For given s, we obtain an explicit estimator. By standard influence techniques we obtain

$$var\,[\hat{\lambda}_{cf}(s)] \approx n^{-1}2^{-1}(s+s^{-1})^2 e^{s^2}\,[1+s^2-e^{-2s^2}(1+3s^2)/(1+4s^2)]\,. \tag{4.61}$$

It is clear from a graph of this function that, to minimize variance, we should choose $s = 0$. For this choice of s,

$$\begin{aligned}\hat{\lambda}_{cf} &= \lim_{s\to 0} s\hat{R}\,\hat{I}^{-1} \\ &= \lim_{s\to 0} \sum_{j=1}^{n} \cos\,(sZ_j) / \sum_{j=1}^{n} \{\,\sin\,(sZ_j)\,Z_j/(sZ_j)\,\} \\ &= \bar{Z}^{-1}\,.\end{aligned} \tag{4.62}$$

So this is an example of a degenerate transform estimator, as discussed in Example 3.6. To obtain a robust estimator we must choose $s > 0$, and for our present purposes it is sufficient to set $s = 1$, enabling us to compare the m.l.e. with an IBR alternative.

The following table demonstrates the practical significance of our earlier influence results. This simulation study was based on 1000 samples, each of size 100, for $\lambda = 1$. Note that the "E" notation in the body of the table is used to signify powers of 10.

It is clear from this simulation study that the m.l.e. is of little practical use in the face of contaminated data. The evidence suggests that one should employ a robust alternative, especially if λ is a parameter of interest. This is another argument for future work to focus on ISE methods, which is the subject of the next section.

Table 4.3. Simulation Results for $\hat{\lambda}_{ml}$ versus $\hat{\lambda}_{cf}$.

ε	ML		CF	
	Bias	MSE	Bias	MSE
0	2.29E-3	1.78E-3	-1.436E-3	1.05E-2
0.1	0.4276	0.2064	-2.762E-3	1.313E-2
0.2	0.4836	0.2403	1.633E-2	6.548E-4

5. Influence for the ISE-Estimator.

The ISE function is

$$h(\Theta) = \int_{-\infty}^{\infty} |\hat{\phi}(s) - \phi(s;\Theta)|^2 e^{-l^2 s^2} ds\,, \tag{4.63}$$

where $\Theta = (r, \lambda, \mu, \sigma^2)'$ and we employ the same form for the weight-function as in Chapter 3. Note that the free parameter is denoted l, rather than λ as before, for obvious notational reasons. We may write the characteristic function as

$$\phi(s;\Theta) = \lambda^r (\lambda^2 + s^2)^{-r/2} e^{-s^2\sigma^2/2} e^{i(s\mu + r\tan^{-1} s/\lambda)}\,, \tag{4.64}$$

allowing easy access to the real and imaginary parts. Given a random sample $Z_1, \ldots, Z_n$, we may estimate this quantity as

$$\hat{\phi}(s) = n^{-1} \sum_{j=1}^{n} e^{isZ_j}\,.$$

Recalling Theorem 3.1 we have, for the ISE-estimator $\hat{\Theta}$ of Θ,

$$IF(X;\hat{\Theta}) = \kappa^{-1}\underline{\tau}\,, \tag{4.65}$$

where

$$(\tau)_j = \int_{-\infty}^{\infty} [\frac{\partial R}{\partial \theta_j} IF(X;\hat{R}) + \frac{\partial I}{\partial \theta_j} IF(X;\hat{I})] e^{-l^2 s^2} ds \quad j=1,..,4. \tag{4.66}$$

To aid the discussion to follow, we write the above expression as

$$(\tau)_j = \int_{-\infty}^{\infty} \psi_j(s) e^{-l^2 s^2} ds. \tag{4.67}$$

The elements of the matrix κ may be expressed in similar fashion as

$$(\kappa)_{ij} = \int_{-\infty}^{\infty} \omega_{ij}(s) e^{-l^2 s^2} ds \quad i,j=1,..,4, \tag{4.68}$$

where

$$\omega_{ij}(s) = \frac{\partial R}{\partial \theta_j}\frac{\partial R}{\partial \theta_i} + \frac{\partial I}{\partial \theta_j}\frac{\partial I}{\partial \theta_i}.$$

It was not possible to evaluate the elements of $\underline{\tau}$ and κ explicitly, so a Gauss-Hermite quadrature was employed to obtain approximate solutions. Thus,

$$(\tau)_j \approx \sum_{k=1}^{64} w_k e^{(1-l^2)s_k^2} \psi_j(s_k), \tag{4.69}$$

given a 64-point quadrature. A similar approximation for $(\kappa)_{ij}$ follows from (4.68). Given these quantities, it is straightforward to obtain the influence vector (4.65) using MAPLE. Note that the theoretical work of Chapter 3 tells us that the influence functions so obtained will be bounded, so that the corresponding estimators are IBR. We focus again now on the illustrative example $\Theta = (1,1,0,1)'$, the influence functions for which are depicted in Figures 4.8-4.12 below. The quadrature weights $\{w_k\}$ and sampling points $\{s_k\}$ were obtained from NAG routine D01BCF.

The range of values plotted, $X \in [-10,10]$, was chosen to comfortably enclose the region normally associated with the Lagged-Normal distribution under consideration; the sum of N(0,1) and exp(1) random variables tending to lie between -1 and +5.

For each estimator the free parameter is $l = 1$, with the case $l = 0$ also given for $\hat{r}$. This is in order to illustrate the effect observed in Figure 3.1, where decreasing the free parameter reduced sensitivity to outlying observations. This is still the case, but in a more sophisticated fashion. The effect of reducing l to 0 has been to even-out the peaks and troughs. We do not give the corresponding plots for the other parameters, but this observation holds true for them also. Note that these influence functions, being sinusoidal in nature, eventually repeat themselves.

For moment estimation, the theory developed in Chapter 3, allied with symbolic algebra, enabled relatively straightforward calculation of the influence functions required. It was noted at the end of Theorem 3.1 that the result (4.65) may not provide a particularly efficient means for calculating these for ISE. The reason for this caution is the requirement that a number of compli-

cated integrals be evaluated. The work of this section shows that combining (4.65) with a suitable quadrature rule provides a very efficient route to the desired influence functions.

Figure 4.8. Plot of $IF(X;\hat{r})$, where r is estimated by ISE.

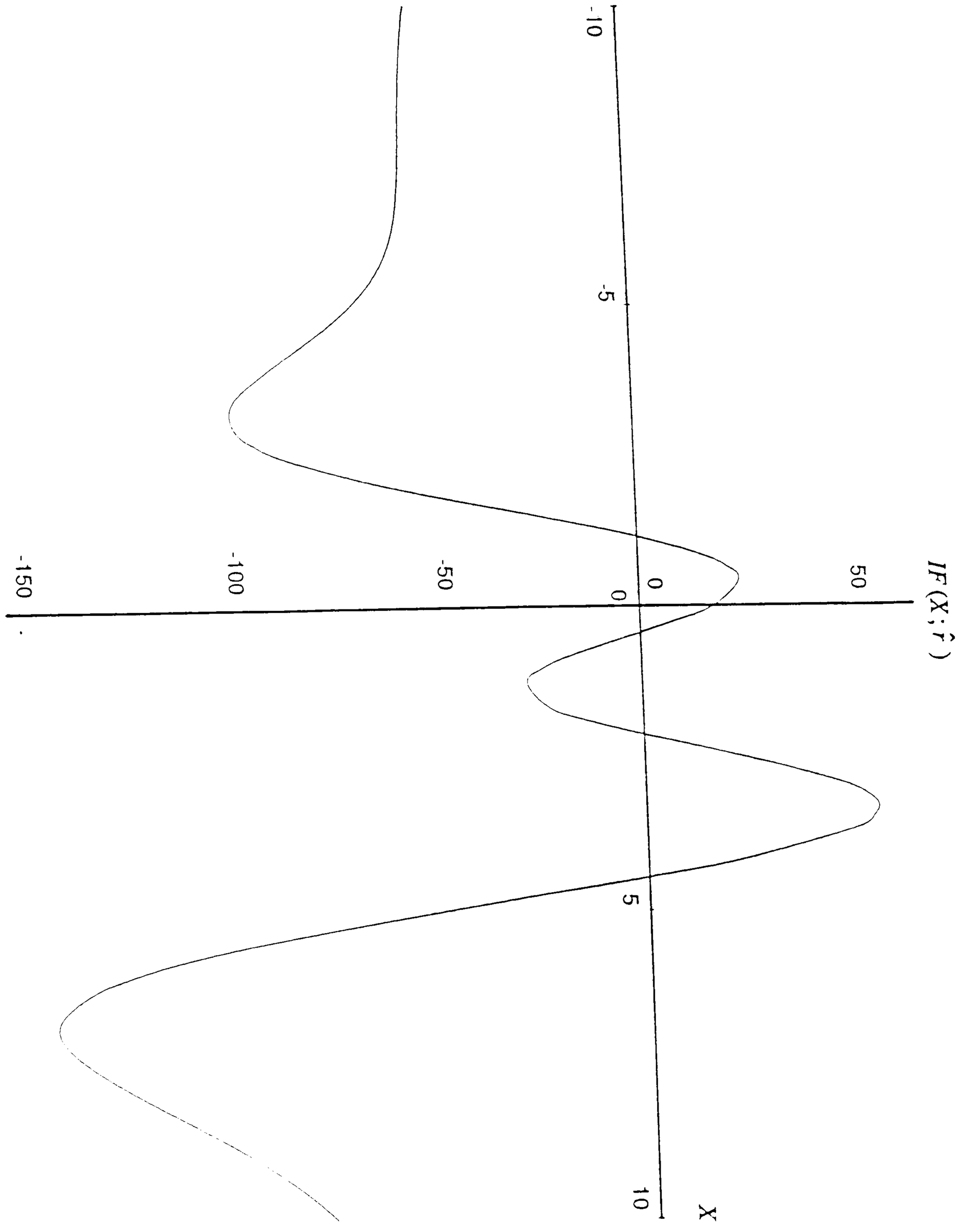


Figure 4.9. Plot of $IF(X;\hat{r})$, with $l = 0$.

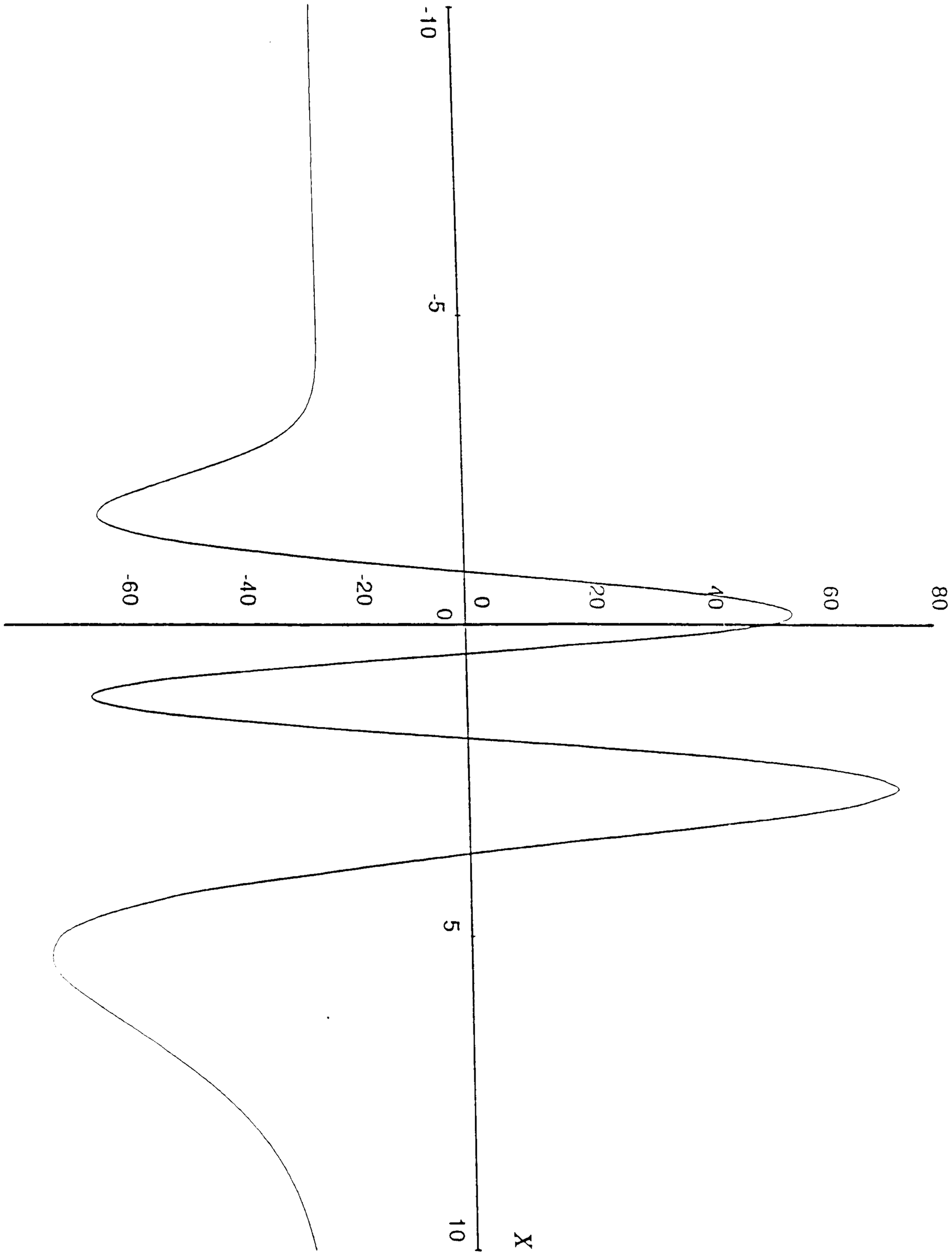


Figure 4.10. Plot of $IF(X;\hat{\lambda})$.

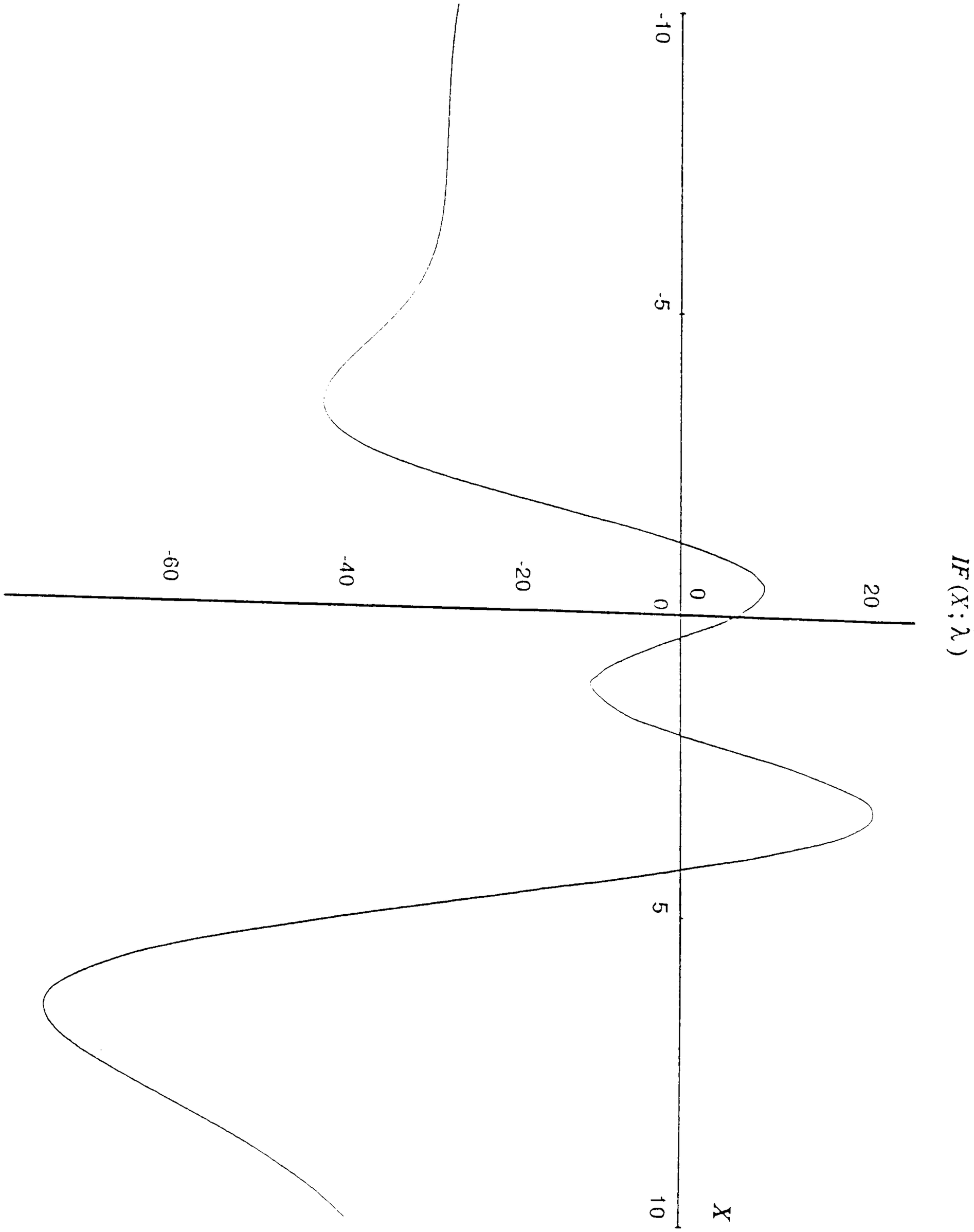


Figure 4.11. Plot of $IF(X;\hat{\mu})$.

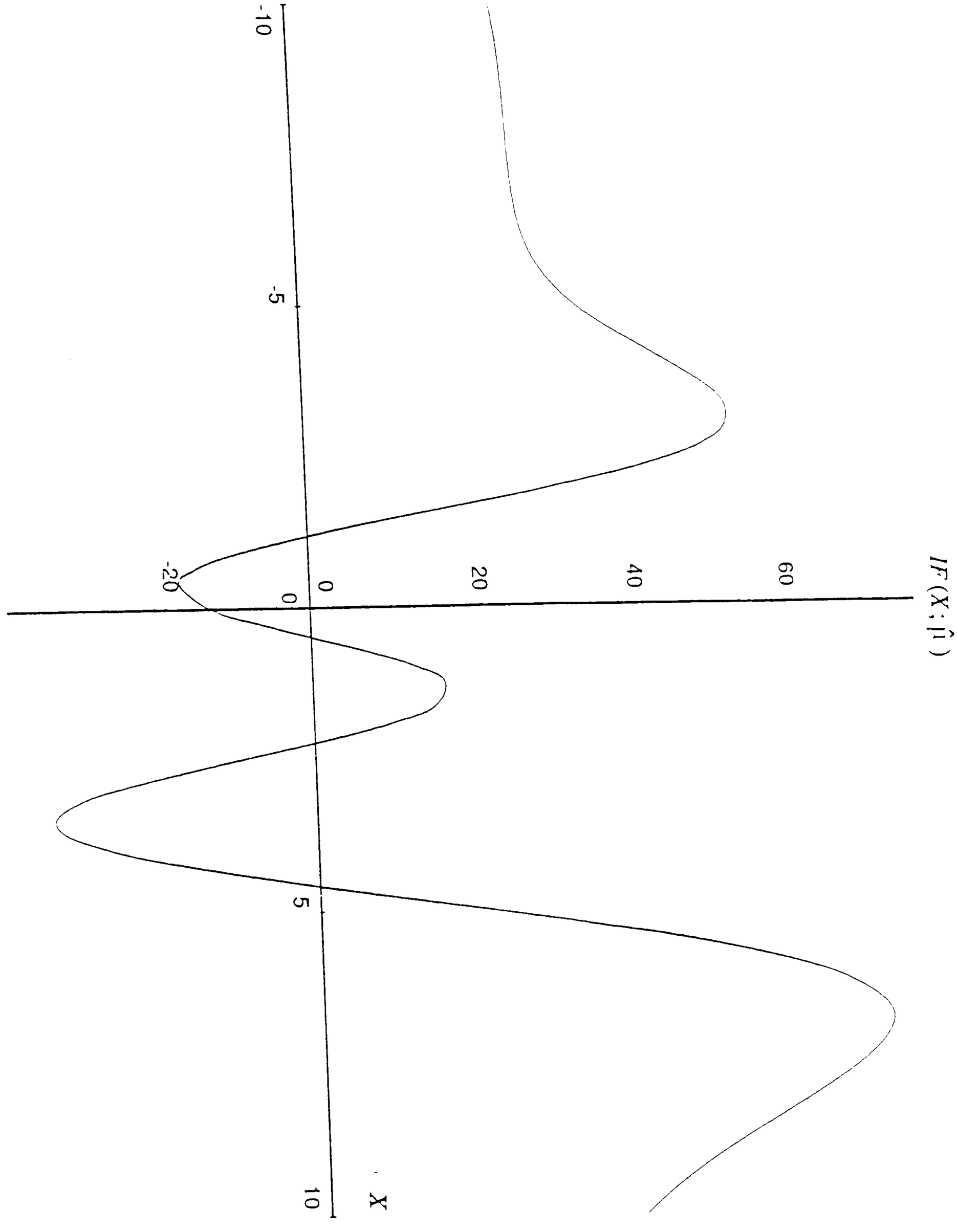
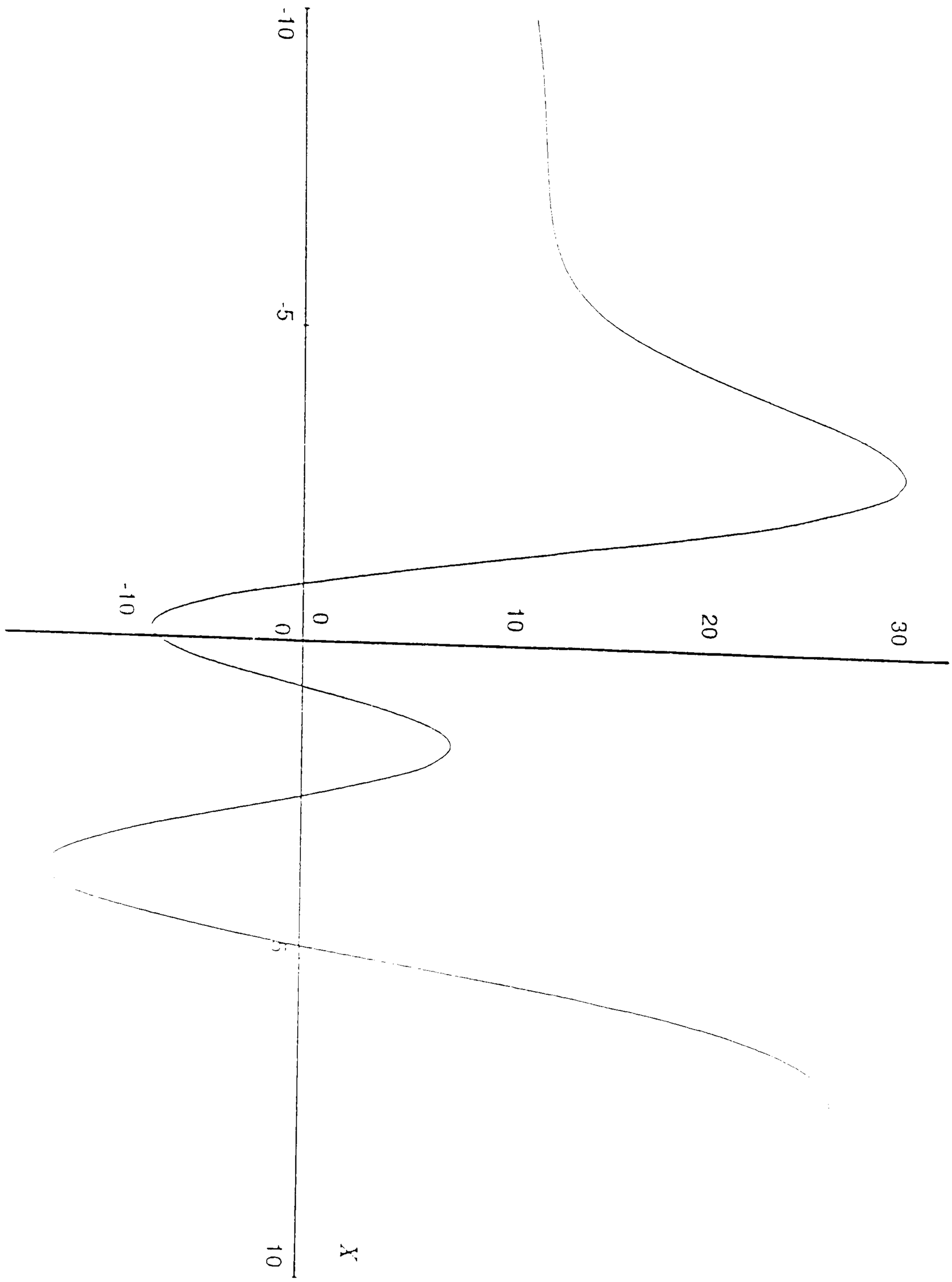


Figure 4.12. Plot of $IF(X;\hat{\sigma^2})$.



CHAPTER 5:

ESTIMATION FOR INDEXED RANDOM VARIABLES.

1. Introduction.

We now turn our attention away from the classical setting of independent and identically distributed random variables, and towards a study of indexed random variables. Typically, although not exclusively, this indexing is with respect to time. In the context of complex stochastic models, explicit parameter estimation is rarely possible, so that one must normally rely on numerical methods. In this chapter we see that techniques based on transforms can yield explicit estimators, whilst offering the possibility of reducing numerical complexities. However, in the context of i.i.d. random variables we saw that transform construction was straightforward. This is not the case here, but we do investigate two competing methods that are reasonably easy to apply in practice.

In section 2 we suggest two forms of stochastic model which are of general application. In the first we consider random variables indexed by the dose of some substance. In the second we look at a very general class of model indexed by time. Important features of the discussion here concern the choice of transform to employ and the quantity to be transformed. Section 3 focuses on the mechanics of transform construction. In particular, one of the techniques is based on Gauss-Laguerre quadrature, and so allows an element of experimental design to take place. Having discussed some of the fundamentals of transform methods in this context, we move on in section 4 to illustrate these methods by applying them to a Quantal Assay model. We are then in a position to focus on more complex stochastic models. As a vehicle to facilitate this discussion we employ the One-Compartment Open model; although it is a very simple member of the general class of compartment models, it has found important practical applications. We see that explicit parameter estimators, based on the method-of-moments, are available. Two crucial issues arise: we must be able to choose a value for the transform variable in order to produce explicit estimates; secondly, we can only hope to produce reliable parameter estimates if the empirical

transform is itself a reliable estimator. These issues are explored before we move on to consider the performance of the estimators themselves.

A key feature of transform estimators is that we need only employ a single value of s, as shown in Appendix II. This approach did not work well in Chapter 4, but was based there on a modified transform. As a consequence of this, we were having to estimate high-order derivatives of the moment generating function. We aim to avoid this here by employing an un-modified transform, offering the possibility of considerably reducing the dimension of search required to select the transform variable. The work of this chapter is completed by considering some more general compartment models where a transform approach may be of practical use.

2. The Transform Approach to Fitting Stochastic Models.

In the work to date we have used transforms which are functionals of some c.d.f. F. These were either of the form (3.1), or modified versions such as those employed in Chapter 4. In the context of stochastic models the quantity to transform is often less clear. We may be guided in this choice by the desire to produce explicit parameter estimators if possible. We must also select the type of transform to employ. In this thesis we only discuss positive indexing, such as by time. Given our previous influence work, the Laplace transform seems a reasonable quantity to employ. Thus, if we wish to transform $\gamma(t;\Theta)$, where t denotes the indexing variable, we form

$$L(s;\Theta) = \int_0^\infty e^{-st}\,\gamma(t;\Theta)\,dt. \tag{5.1}$$

In particular examples we may wish to make use of a different transform, to take account of a particular form of γ for example, but we concentrate on the Laplace transform here.

2.1. A Quantal Assay Model.

In the first application discussed here we consider data from Finney (1971, pp 20) relating to an assay of the chemical rotenone. The table below lists the log-dose (base 10), x, of rotenone in mg/litre. The chemical is applied to aphids, with response being death or seriously affected. For each log-dose x_j, n_j aphids are tested and r_j respond.

Table 5.1. Results of an Assay of Rotenone.

j	x_j	n_j	r_j
1	0.41	50	6
2	0.58	48	16
3	0.71	46	24
4	0.89	49	42
5	1.01	50	44

We may model these data using a logistic link function as follows.

$$r_j \sim Bin\,(n_j, P\,(x_j))\ ,\ P\,(x) = [1+e^{-(\alpha+\beta x)}]^{-1}\ . \tag{5.2}$$

The details of fitting this model by maximum likelihood, using BMDP, are summarized below.

Table 5.2. Maximum Likelihood Fit of Quantal Assay Model.

Parameter	Estimate	S.E.	Estimate/ S.E.
INTERCEPT	-4.839	0.638	-7.59
SLOPE	7.068	0.883	8.01
log-likelihood ($\hat{\alpha}$, $\hat{\beta}$) = -119.856			

The logit model provides a very good fit to these data.

Laurence et al. (1986) note that $[1-P\,(x)]^{-1} = 1+e^{\,\alpha+\beta x}$, so we may exploit this simple form in the parameters and employ

$$L(s\,;\,\Theta\,) = \int_0^\infty e^{-sx}[1+e^{\,\alpha+\beta x}]dx\ , \tag{5.3}$$

where $\Theta = (\alpha,\beta)'$. We therefore find

$$L\,(s\,;\,\Theta\,) = s^{-1}+e^{\,\alpha}(s-\beta)^{-1} \quad s>\beta\ . \tag{5.4}$$

This example is returned to later in this chapter when we have developed transform methods a little further.

2.2. Models of the Form $Y(t) = \mu(t; \Theta) + \varepsilon(t)$.

The obvious choice here is $\gamma = \mu$, so that

$$L(s; \Theta) = \int_0^\infty e^{-st} \mu(t; \Theta) dt . \tag{5.5}$$

Later in this chapter we focus on the One-Compartment Open model, which may be expressed in this form. Under this scenario we observe the random variables $\{Y(t_1),...,Y(t_n)\}$, for a grid of sampling points $\underline{t} = (t_1, \ldots, t_n)'$ to be determined.

In order to estimate the parameters of the models we have discussed in section 2, we form the estimating equations

$$\hat{L}(s_i) = L(s_i; \hat{\Theta}) \quad i=1,...,p. \tag{5.6}$$

Alternatively, we may allow the $\{s_i\}$ to tend to some common value, s say, yielding the estimating equations

$$\hat{L}^{(i-1)}(s) = L^{(i-1)}(s; \hat{\Theta}) \quad i=1,...,p. \tag{5.7}$$

For details, see Appendix II. Both of these techniques will be investigated extensively during this chapter. We now move on to consider the key issue of transform construction.

3. Transform Construction.

Two distinct approaches are considered here. We first look into methods based on numerical quadrature, where we find that an element of experimental design is possible. The second technique follows from a Riemann-sum approximation to the integral (5.1), which has the potential disadvantage of not providing any objective means of choosing the sampling points. However, this technique is suitable for situations where no such experimental design has been possible.

3.1. Methods Based on Quadrature.

We note first that the Laplace transform (5.5) may be re-expressed as

$$L(s; \Theta) = \int_0^\infty e^{-t} f(t) dt , \tag{5.8}$$

where

$$f(t) = \mu(t; \Theta)e^{t(1-s)} . \tag{5.9}$$

The objective of numerical quadrature is to form an estimate of $L(s; \Theta)$ as a weighted-sum of values of $f(t)$. In our statistical application these are unknown. We may, however, estimate $f(t) = \mu(t; \Theta)e^{(1-s)t}$ by $Y(t)e^{(1-s)t}$. We therefore obtain the empirical transform

$$\hat{L}(s) = \sum_{k=1}^{n} w_k Y(t_k) e^{(1-s)t_k} , \tag{5.10}$$

for a set of weights $\underline{w} = (w_1, \ldots, w_n)'$ and sampling points $\underline{t}$. The usual approach to choosing $\underline{w}$ and/or $\underline{t}$ is to ensure that $\hat{L}(s) = L(s; \Theta)$ when $f(t)$ is a polynomial of suitable order. A similar approach to estimating the more general integral (5.1) is possible, but the approximation of γ is clearly particular to the problem concerned.

Returning to the problem at hand, there are two cases to consider. If we are at the stage of planning an experiment, then we are able to determine both $\underline{w}$ and $\underline{t}$. This scenario is discussed next. Following this we discuss the situation where data have been collected at arbitrary sampling points, so that only the weights remain to be determined. It is shown later, in accordance with common sense, that we can achieve far superior results from designed experiments. We complete section 3 by suggesting that we are not only restricted to polynomial rules, that is, when $f(t)$ is assumed to be at least well-approximated by a finite polynomial, and indicate some possible ways forward.

3.1.1. Data Collected at Arbitrary Sampling Points.

We have only to determine the n weights to form $\hat{L}(s)$. Thus the quadrature rule is exact for polynomials of degree $n-1$ at most, and is said to be a rule of precision $n-1$ as a result. The weights are found as follows, via the method discussed by, for example, Philips & Cornelius (1986) pp 273-74.

$f(t) = 1$

$$\hat{L}(s) = L(s) \Rightarrow \sum_k w_k = \int_0^\infty e^{-t} dt = \Gamma(1)$$

$f(t) = t$

$$\hat{L}(s) = L(s) \Rightarrow \sum_k w_k t_k = \int_0^\infty te^{-t}dt = \Gamma(2),$$

so that the jth equation is

$$\sum_{k=1}^{n} w_k t_k^{\,j-1} = \Gamma(j) \quad j=1,...,n\ . \tag{5.11}$$

We therefore obtain the weights from

$$T\underline{w} = \underline{\Gamma}\ , \tag{5.12}$$

where

$$T = \begin{pmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_n \\ . & . & . \\ . & . & . \\ t_1^{\,j-1} & t_2^{\,j-1} & t_n^{\,j-1} \\ . & . & . \\ . & . & . \\ t_1^{\,n-1} & t_2^{\,n-1} & t_n^{\,n-1} \end{pmatrix}$$

and

$$\underline{\Gamma} = (\Gamma(1),\Gamma(2),...,\Gamma(n))'\ .$$

We find, then, that

$$\underline{w} = T^{-1}\underline{\Gamma}\ . \tag{5.13}$$

Note that, provided $t_m \neq t_n \forall\ m \neq n$, then T^{-1} is defined. Given this condition, which is clearly satisfied for any practical example, the weights are well-defined. In the following example we use the Quantal Assay model to illustrate the above quadrature rule.

Example 5.1.

We return to the quantal assay example discussed earlier. The empirical transform may be written as

$$\hat{L}(s) = \sum_{j=1}^{5} w_j e^{(1-s)x_j}[1-\hat{P}(x_j)]^{-1}\ ,$$

where $\hat{P}(x_j)$ denotes the empirical proportion responding at x_j. Note that problems arise if $\hat{P}(x_j) = 1$, but this does not happen here. The form of this estimator follows from (5.3) and applying (5.10) to this slightly different example. We therefore find

$$\hat{L}(s) = \sum_{j=1}^{5} w_j e^{(1-s)x_j} [1 - r_j/n_j]^{-1} ,$$

a quadrature rule of precision 4, with the weights found as follows.

$$T = \begin{bmatrix} 1 & 1 & . & 1 \\ 0.41 & 0.58 & . & 1.01 \\ 0.41^2 & 0.58^2 & . & 1.01^2 \\ . & . & . & . \\ 0.41^4 & 0.58^4 & . & 1.01^4 \end{bmatrix}$$

Then

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ . \\ w_5 \end{bmatrix} = T^{-1} \begin{bmatrix} 1 \\ 1 \\ 2 \\ 6 \\ 24 \end{bmatrix} .$$

The symbolic algebra package MAPLE was used to solve for $\underline{w}$, exploiting the special structure of T, yielding the vector of weights

$$\underline{w} = \begin{bmatrix} 736.0815497 \\ -3807.9016440 \\ 5492.605598 \\ -3766.857507 \\ 1347.072009 \end{bmatrix} .$$

It is straightforward to verify that this quadrature scheme is exact when $f(x)$, in the notation of equation (5.8), is a polynomial of degree at most 4.

There are some points worthy of note arising from the above calculation. Firstly, these weights are, in modulus, very large. This could well lead to $\hat{L}(s)$ being numerically unstable. Second, a larger Quantal Assay data set was investigated, and the weights became even larger in absolute terms. It seems that this quadrature technique may be difficult to apply in practice.

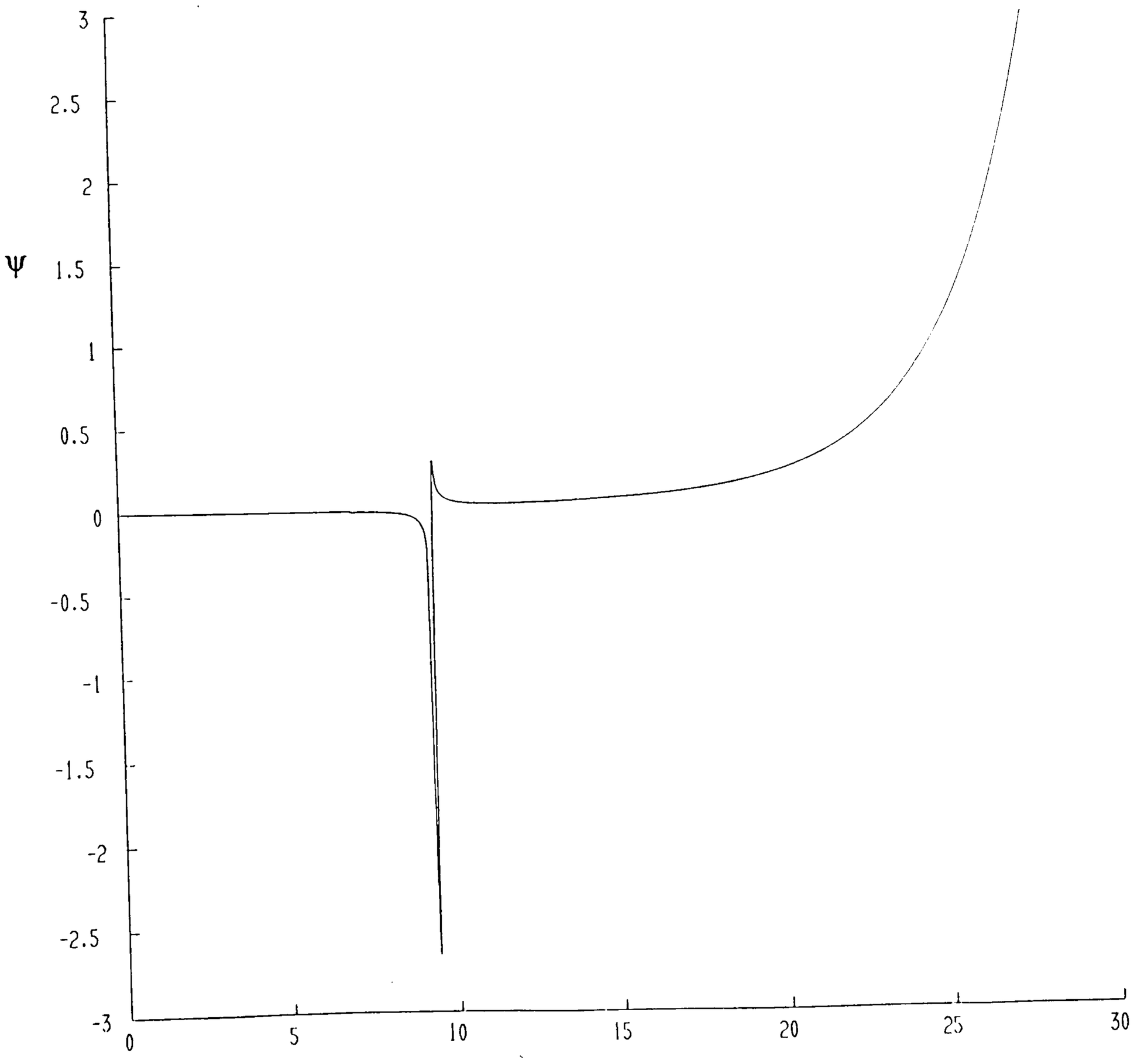
In order to investigate the performance of $\hat{L}(s)$ for this example, we define the ratio

$$\psi(s) = L(s;\, \alpha = -4.839,\, \beta = 7.068)/\hat{L}(s)\,,$$
$$= L_m(s)/\hat{L}(s) \quad say,$$

where we substitute the m.l.e.s of α and β in the Laplace transform to enable a comparison to be made. A plot of $\psi(s)$ follows below, and we look for regions where this ratio is close to unity. It is clear from this that $\hat{L}(s)$ does not perform very well at all; this is perhaps unsurprising given that we only achieve precision 4, as well as the instability of $\underline{w}$ discussed above.

As we see in the next sub-section it is possible to obtain a scheme of precision $2n-1$. The practical implications are investigated later, but one would anticipate a noticeable improvement in the performance of the empirical transform with this substantial increase in precision. Had the Quantal Assay experiment of Example 5.1 been planned with quadrature in mind, we could then have achieved precision 9.

Figure 5.1. Plot of $L_m(s)/\hat{L}(s)$.



3.1.2. Designed Sampling Points.

The assumption here is that the data are yet to be collected. We are therefore able to determine not only the weights $\underline{w}$, but the sampling points $\underline{t}$ as well. To do this, we construct a system of $2n$ equations in exactly the same manner as before, to give

$$\sum_{k=1}^{n} w_k t_k^{\,j-1} = \Gamma(j) \quad j=1,\ldots,2n \ . \tag{5.14}$$

The sampling points $\underline{t} = (t_1, \ldots, t_n)'$ are the roots of the nth Laguerre polynomial $L_n(t)$, whilst the weights are found as

$$w_k = t_k [L_{n+1}(t_k)]^{-2} \ . \tag{5.15}$$

Such a quadrature rule is known as a Gauss-Laguerre scheme of precision $2n-1$.

Fortunately, the sampling points and corresponding weights are widely tabulated for a variety of sample sizes. See, for example, Abramowitz & Stegun (1970) pp 923. They are also readily available via the NAG FORTRAN Library; NAG routine D01BBF was employed in this thesis. For general references to Gaussian quadrature, Philips & Cornelius (1986) is an excellent practical guide. For more theoretical detail one might consult the standard reference Rabinowitz & Davis (1984).

The quadrature rules we have discussed so far are based on a polynomial approximation. In the next section we suggest two alternatives to this, although we do not pursue them here.

3.1.3. Alternative Approaches.

The techniques suggested here assume that $\underline{t}$ has already been chosen, although it does seem that they could be extended to encompass this choice as well.

The first suggestion we discuss is based on a least-squares criterion. We still form the empirical transform as a weighted-sum, but choose $\underline{w}$ by minimizing

$$S(\underline{w}) = \int_0^{\infty} [\hat{L}(s) - L(s;\Theta)]^2 ds \ . \tag{5.16}$$

Now,

$$S(\underline{w}) = \int_0^{\infty} [\sum_k w_k f(t_k) - L(s;\Theta)]^2 ds$$

$$\Rightarrow \quad \frac{\partial S}{\partial w_j} = 2\int_0^\infty f(t_j)[\sum_k w_k f(t_k) - L(s;\Theta)]ds$$

$$\Rightarrow \quad \frac{\partial^2 S}{\partial w_j^2} = 2\int_0^\infty [f(t_j)]^2 ds > 0 \quad \forall\, \underline{w} \in R^n .$$

Thus the solution of the system

$$\frac{\partial S}{\partial w_j} = 0 \qquad j=1,\ldots,n$$

is guaranteed to provide a minimum.

For the remainder of this discussion we substitute the explicit form

$$f(t_k) = Y(t_k)e^{(1-s)t_k} ,$$

so that $\underline{w}$ results as the solution of

$$\int_0^\infty e^{(1-s)t_j}[\sum_k w_k Y(t_k)e^{(1-s)t_k} - L(s;\Theta)]ds = 0 . \tag{5.17}$$

On rearranging the above expression a little, we find

$$\sum_k w_k Y(t_k)\int_0^\infty e^{(1-s)(t_j+t_k)}ds = \int_0^\infty e^{(1-s)t_j}L(s;\Theta)ds .$$

Defining the matrix κ having kj-th element

$$(\kappa)_{kj} = Y(t_k)\int_0^\infty e^{(1-s)(t_j+t_k)}ds \qquad k,j=1,\ldots,n \tag{5.18}$$

and the vector $\underline{l}(\Theta)$ having jth element

$$[\underline{l}(\Theta)]_j = \int_0^\infty e^{(1-s)t_j}L(s;\Theta)ds \qquad j=1,\ldots,n, \tag{5.19}$$

then the least-squares solution for $\underline{w}$ satisfies

$$\kappa\underline{w} = \underline{l}(\Theta)$$

$$\Rightarrow \quad \underline{w} = \kappa^{-1}\underline{l}(\Theta) . \tag{5.20}$$

The drawback of this approach is that the weights depend explicitly on the unknown parameters. An adaptive procedure would be one way forward, but this does seem unnecessarily complicated. Alternatively, we might investigate the sensitivity of $\underline{w}$ to the particular choice of Θ. Were it the case that the elements of $\underline{w}$ are largely insensitive to changes in Θ, then we could justify using a particular least-squares solution. However, this seems rather speculative at this point.

To conclude this discussion, note that a least-squares approach was adopted by Katz & D'Argenio (1983), concentrating on the choice of the sampling points $\underline{t}$. However, they only obtain the optimal sampling points for particular choices of Θ.

The second suggestion is to choose $\underline{w}$ so that the quadrature rule is exact when $f(t)$ is a mixture of exponential terms. The key to this approach is the choice of a suitable set of linearly independent functions. For the polynomial rules we employed the set $\{1, t, t^2, \ldots, t^{n-1}\}$; here we choose $E = \{e^{-t\,j/n} : j=1,\ldots,n\}$. We find the quadrature rule which is exact for linear combinations of members of E as follows.

The weights must satisfy the system of equations

$$\sum_{k=1}^{n} w_k e^{-jt_k/n} = \int_0^{\infty} e^{-t}\, e^{-tj/n} dt = n(n+j)^{-1} \qquad j=1,\ldots,n. \tag{5.21}$$

In the notation established earlier, the matrix T has elements

$$(T)_{ji} = e^{-t_i\, j/n},$$

so that

$$n^{-1} T\underline{w} = \begin{bmatrix} (n+1)^{-1} \\ (n+2)^{-1} \\ . \\ . \\ (2n)^{-1} \end{bmatrix}. \tag{5.22}$$

Given the way E has been selected, T is clearly non-singular. The method is, therefore, well-defined. Its efficiency will depend on how well $f(t)$ may be approximated by linear combinations of the elements of E. We note in passing that the approach considered here is not unique, since other choices of E are clearly possible. The particular form for E would be chosen to exploit some feature of the application under consideration.

3.2. A Riemann-Sum Approach.

In the notation of equation (5.1), we have the problem of calculating the area under the curve $e^{-st}\,\gamma(t;\Theta)$, for t between 0 and ∞. We may do this by approximating the area by rectangles. Now,

$$L(s;\Theta) = \sum_{i=1}^{n} \int_{c_{i-1}}^{c_i} e^{-st}\,\gamma(t;\Theta)\,dt\ , \tag{5.23}$$

where

$$c_0 = 0\ ,\ c_n = \infty\ ,\ \{c_j : 1 \le j \le n-1\}$$

are constants chosen so that $c_{j-1} \le t_j \le c_j$, $1 \le j \le n$. Although not a unique choice, we employ

$$c_j = 0.5\,(t_j + t_{j+1}) \quad 1 \le j \le n-1\ , \tag{5.24}$$

so that the c_js lie half-way between the sampling points. We may now make the Riemann-Sum approximation, viz:

$$L(s;\Theta) \approx \sum_{i=1}^{n} \gamma(t_i;\Theta) \int_{c_{i-1}}^{c_i} e^{-st}\,dt$$

$$= s^{-1} \sum_{i=1}^{n} \gamma(t_i;\Theta)(e^{-sc_{i-1}} - e^{sc_i})\ .$$

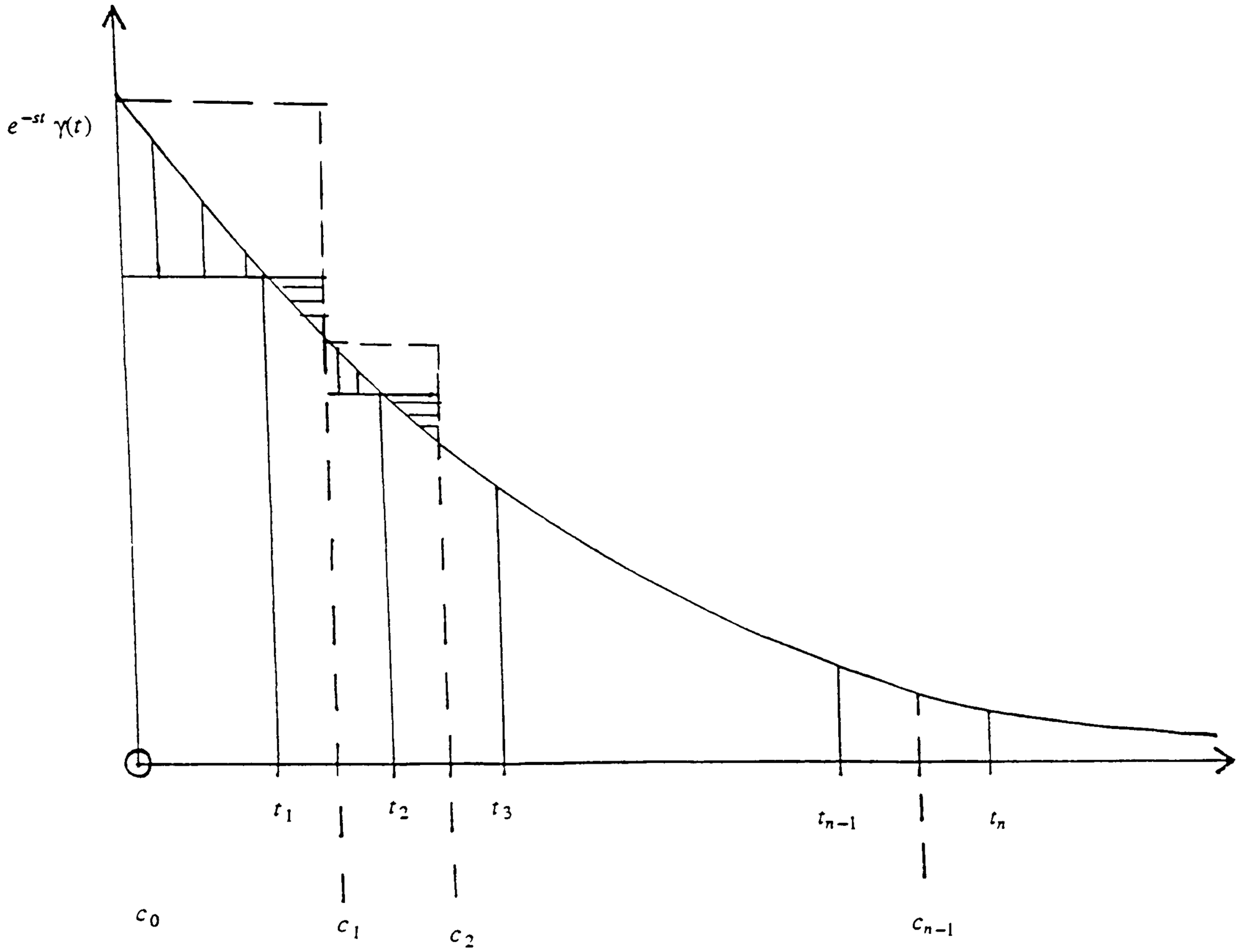
It remains only to estimate γ from the observations $\{\,Y(t) : t \in T\}$, for some indexing set T. Denoting this estimate by $\hat{\gamma}$, we obtain the empirical transform

$$\hat{L}(s) = s^{-1} \sum_{i=1}^{n} \hat{\gamma}(t_i)(e^{-sc_{i-1}} - e^{-sc_i})\ . \tag{5.25}$$

For models of section 5.2.2, we have $\hat{\gamma}(t_i) = Y(t_i)$. In the Quantal Assay model the form of $\hat{\gamma}$ is less straightforward, but follows readily due to the simple interpretation for γ in that context.

This form of transform estimator has been employed by a number of researchers, including, for example, Laurence et al. (1986), Leedow & Tweedie (1983) and Schuh & Tweedie (1979). The Riemann-Sum approximation is depicted below in Figure 5.2. This method will work best when the shaded areas within each rectangle cancel each other out. A corollary of this is that we should exercise caution in regions where the integrand is strongly curved. For Laplace transforms this will tend to occur, if at all, towards the origin. Figure 5.2 shows that this is a particular concern when t_1 is far from 0. This estimator is open to a further criticism in that the range of integration extends beyond the range of the data. Solutions to these problems have been suggested, and are the topic of the next section. We should also note that a second approximation has been made, this being that of γ by $\hat{\gamma}$. Clearly the nature of this approximation is example-

Figure 5.2. Riemann-Sum Approximation.



dependent. We go on now to apply this method of transform construction to the Quantal Assay model.

Example 5.1 - continued.

Recalling section 5.2.2, our task here is to estimate

$$L(s;\Theta) = \int_0^\infty e^{-sx}[1-P(x)]^{-1}\,dx.$$

We first calculate the values of the $\{c_j\}$, which are quoted in the table below.

Table 5.3. Values of $\{c_j\}$ for Riemann-Sum Approximation.

j	c_j
0	0.000
1	0.495
2	0.645
3	0.800
4	0.950
5	∞

We may therefore write

$$L(s;\Theta) \approx \sum_{j=1}^{5}[1-P(x_j)]^{-1}\int_{c_{j-1}}^{c_j} e^{-sx}dx$$

$$= s^{-1}\sum_{j=1}^{5}[1-P(x_j)]^{-1}(e^{-sc_{j-1}} - e^{-sc_j}),$$

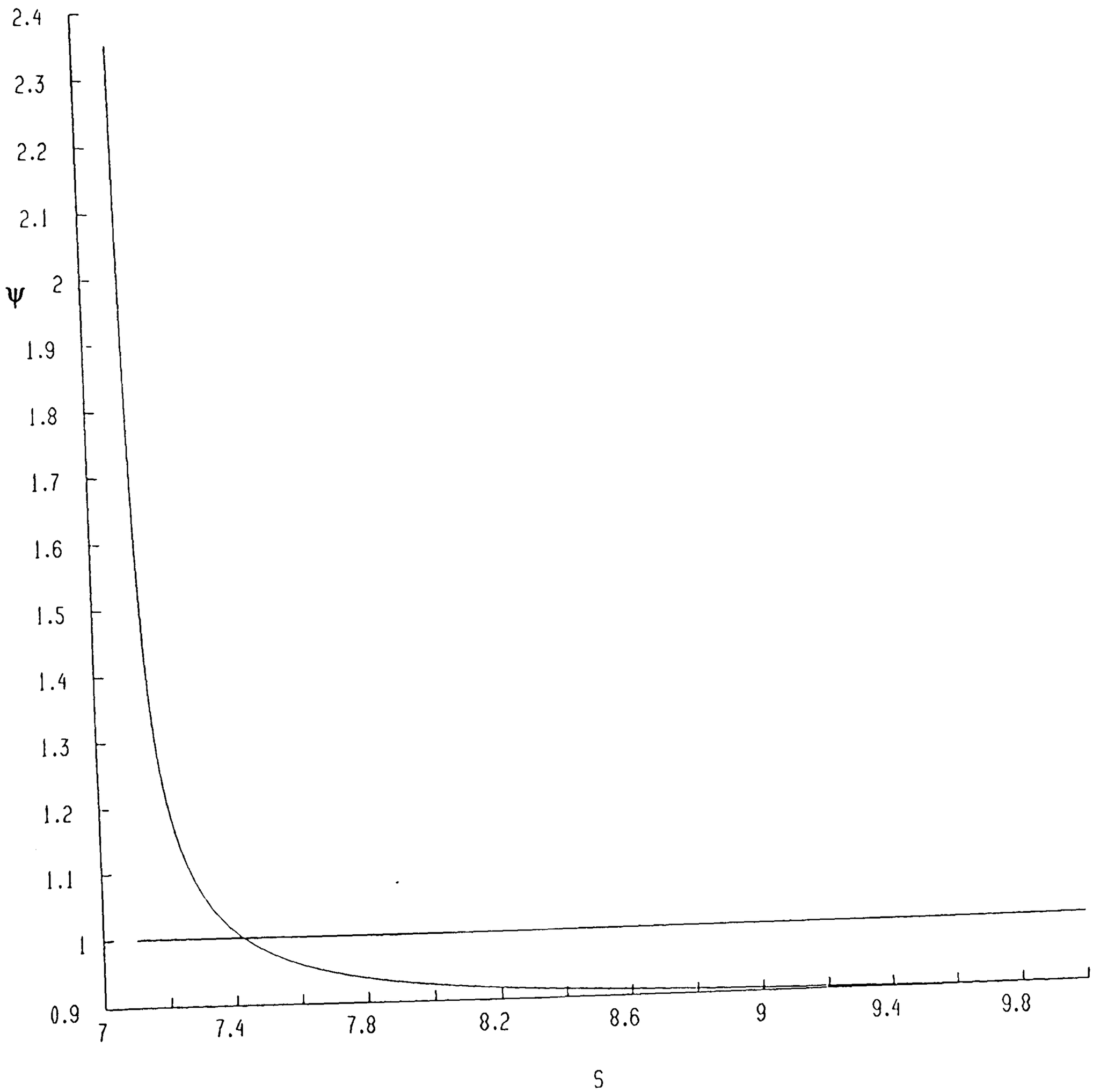
paving the way for the empirical Laplace transform (ELT)

$$\hat{L}(s) = s^{-1}\sum_{j=1}^{5}[1-r_j/n_j]^{-1}(e^{-sc_{j-1}} - e^{-sc_j}).$$

We again (see section 3.1.1) investigate the performance of the ELT via the ratio $L_m(s)/\hat{L}(s)$, depicted in Figure 5.3 below. We only plot that part which is $> \hat{\beta}_{ML}$. This transform estimator clearly performs much better than the one based on quadrature. This is perhaps a false

comparison though, since the Riemann-Sum estimator was designed with this application in mind. On a slightly negative note, in the region where agreement is best, $\hat{L}(s)$ consistently overestimates $L(s;\Theta)$. Recalling the general forms (5.10) and (5.25) for the transform estimators, we see that they are clearly biased for any finite sample.

Figure 5.3. $L_m(s)/\hat{L}(s)$ for the Riemann-Sum ELT.



3.2.1. Improvements to the Riemann-Sum Estimator.

This estimator can be improved in two ways, in response to the potential difficulties identified above. The first problem arises when t_1 is too far from the origin. We may overcome this by a judicious translation of the data, and work instead with

$$y_j = t_j - \tau \quad j=1,...,5, \tag{5.26}$$

for τ to be determined. Note that we retain our general notation for this discussion.

The second problem arises from integrating beyond the range of the data. This is perhaps unwise in itself, but in setting $c_0 = 0$ and/or $c_n = \infty$, we also risk placing too much weight on $Y(t_1)$ and/or $Y(t_n)$. This may well lead to a poor approximation. An approach known as "end-corrections" was suggested by Leedow & Tweedie (1983) to avoid these detractions. We concentrate here on upper end-corrections since lower end-correction may be achieved by translation.

Now, if we set $c_n = t_n$ then

$$\hat{L}(s) = s^{-1} \sum_{j=1}^{n} \hat{\gamma}(t_j)(e^{-sc_{j-1}} - e^{-sc_j})$$

estimates

$$\int_0^{t_n} e^{-st} \gamma(t;\Theta)dt \ .$$

In order to estimate $L(s;\Theta)$ we need also to calculate

$$I(s;\Theta) = \int_{t_n}^{\infty} e^{-st} \gamma(t;\Theta)dt \ . \tag{5.27}$$

For the purposes of parameter estimation we then set

$$\hat{L}(s) + I(s;\Theta) = L(s;\Theta) \ . \tag{5.28}$$

This is not the only approach we might consider. An alternative is, upon setting $c_n = y_n$, to equate the corrected transform directly with its theoretical counterpart. Parameter estimation is then based on

$$\hat{L}(s) = \int_0^{y_n} e^{-sy} [1-P(y)]^{-1} dy$$

$$= s^{-1}(1-e^{-sy_n}) + e^{\alpha}(s-\beta)^{-1}(1-e^{-(s-\beta)}) \ .$$

The complication here is due to the $e^{-(s-\beta)}$ term, rendering explicit estimators unattainable. We do not, therefore, pursue this approach any further. We instead investigate the procedure described by equation (5.28) by means of the following

Example 5.1 - continued.

As suggested above, the ELT can be improved in two ways. First, x_1 is, in relative terms, some way from the origin. There is, therefore, some risk of placing too much weight at this point. We employ a translation of size $\tau = 0.3$ to deal with this. Secondly, we employ an upper end-correction by setting $c_5 = y_5$. We therefore construct our transform estimator based on the information contained in the following table.

Table 5.4. Data for Corrected Transform Approach.

j	y_j	c_j
0	-	0.000
1	0.11	0.195
2	0.28	0.345
3	0.41	0.500
4	0.59	0.650
5	0.71	0.710

We therefore employ $\hat{L}(s)$ to estimate

$$\int_0^{0.71} e^{-sy}\,[1-P(y)]^{-1}dy\,,$$

so that to estimate $L(s;\Theta)$ we need also to evaluate

$$I(s;\Theta) = \int_{0.71}^{\infty} e^{-sy}\,[1-P(y)]^{-1}dy$$

$$= s^{-1}e^{-0.71s} + e^{\alpha+0.71(\beta-s)}(s-\beta)^{-1} \qquad s>\beta\,.$$

Note that $e^{-(\alpha+\beta y_5)} = 1/P(y_5) - 1$, which we may estimate as $1/\hat{P}(y_5) - 1$. Thus

$$e^{-(\alpha+\beta y_5)} \approx n_5/r_5 - 1,$$

leading to the approximation

$$I(s;\Theta) \approx s^{-1}e^{-0.71s}[1+sr_5(n_5-r_5)^{-1}(s-\beta^{-1})] \quad s>\beta\,.$$

For the purposes of parameter estimation we would then set

$$\hat{L}(s)+s^{-1}e^{-0.71s}[1+sr_5(n_5-r_5)^{-1}(s-\beta)^{-1}] = s^{-1}+e^{\alpha}(s-\beta)^{-1}\,. \tag{5.29}$$

This topic is addressed in the next section.

Our current objective is to investigate the quality of the corrected transform estimator. To this end we set $\kappa(s)$ equal to the left-hand-side of (5.29), and plot the ratio $L_m(s)/\kappa(s)$ which is depicted in Figure 5.4 below. In order to calculate $L_m(s)$, we note that the m.l.e. of β is unaffected by the translation, but $\hat{\alpha}_{ml}$ becomes $\hat{\alpha}_{ml}+0.3\hat{\beta}_{ml}$.

We see that the corrected transform estimator performs much better than its cruder relation. In particular, the region of close agreement between the theoretical transform and its estimator is much larger, and they approach equality at a much shallower angle than for the uncorrected version.

4. Parameter Estimation for the Quantal Assay Model.

We have to consider estimation for each method of transform construction, and the techniques employed differ. For quadrature, we may employ directly the methods outlined in section 2; see equations (5.6) and (5.7) along with the surrounding discussion. For the Riemann-Sum approach, we have already shown that the corrected transform is superior. As such, we base estimation on equation (5.29).

4.1. Quadrature.

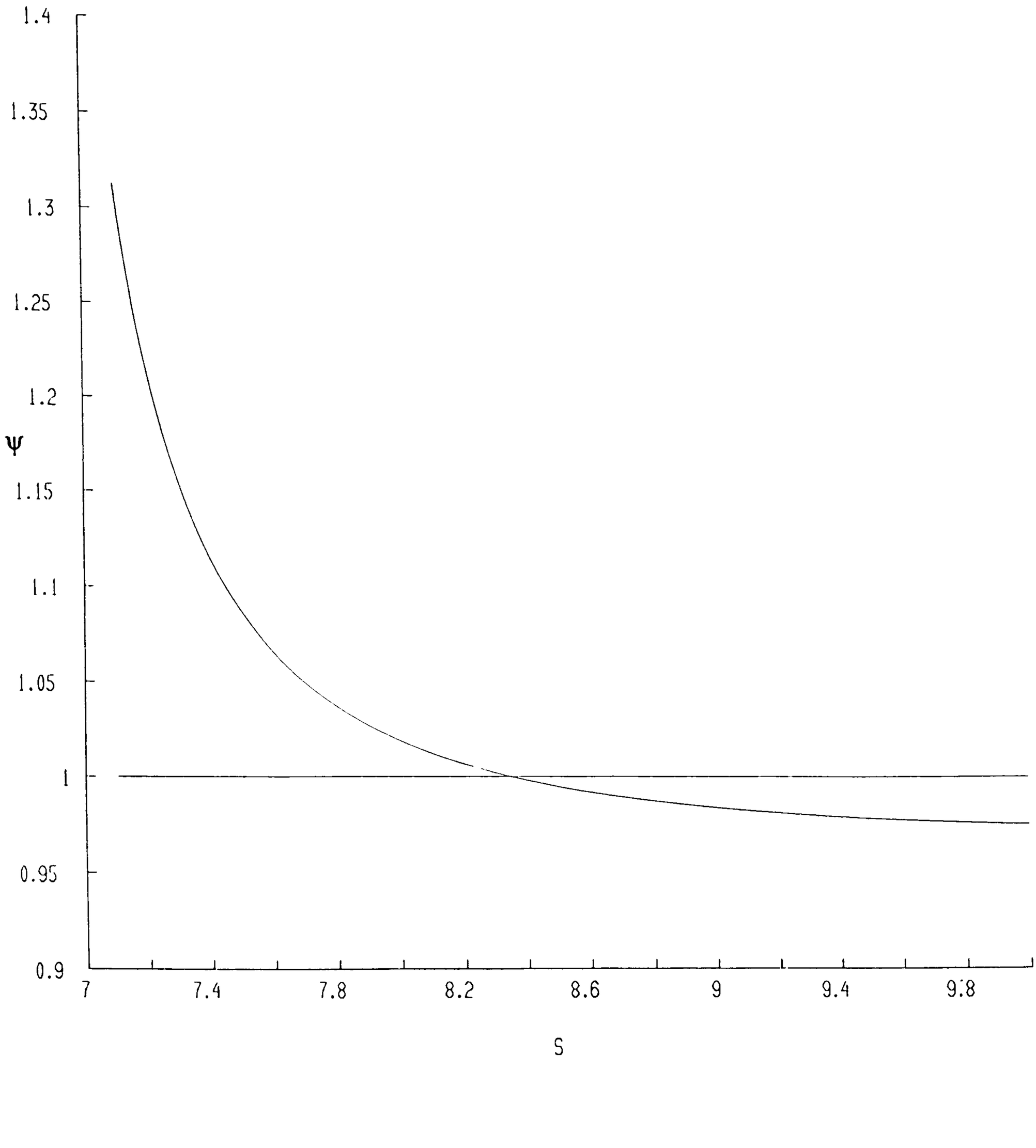
Employing two values of s in (5.6), we obtain the estimating equations

$$\hat{L}_i = s_i^{-1}+e^{\hat{\alpha}}(s_i-\hat{\beta}) \quad i=1,2, \tag{5.30}$$

where L_i denotes $L(s_i)$, as in Chapter 4. Equations (5.30) may be re-cast as

$$s_i\,(s_i\hat{L}_i-1) = (s_i\hat{L}_i-1)\hat{\beta}+s_ie^{\hat{\alpha}} \quad i=1,2.$$

Figure 5.4. Plot of $L_m(s)/\kappa(s)$.



The estimators are found, then, as the solution to the matrix equation

$$\begin{bmatrix} s_1(s_1\hat{L}_1-1) \\ s_2(s_2\hat{L}_2-1) \end{bmatrix} = \begin{bmatrix} s_1\hat{L}_1-1 & s_1 \\ s_2\hat{L}_2-1 & s_2 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ e^{\hat{\alpha}} \end{bmatrix}, \tag{5.31}$$

$$\Rightarrow \quad \begin{bmatrix} \hat{\beta} \\ e^{\hat{\alpha}} \end{bmatrix} = [s_1s_2(\hat{L}_1-\hat{L}_2)+s_1-s_2]^{-1} \begin{bmatrix} s_1s_2(s_1\hat{L}_1-s_2\hat{L}_2) \\ (s_2-s_1)(s_1\hat{L}_1-1)(s_2\hat{L}_2-1) \end{bmatrix}. \tag{5.32}$$

It was found that $\hat{\alpha}(\underline{s})$ does not exist over large regions of $\underline{s}$-space. This is perhaps unsurprising considering how poor the transform estimator is in this context (see Figure 5.1). However, $\hat{L}(s)$ did cross $L(s;\Theta)$ at one point (around $s=25$), so we look next at the estimators which would result from equations (5.7) which employ just one value of s.

The most efficient way to find these estimators is to allow $s_1, s_2 \rightarrow s$ in the explicit forms (5.32). For convenience, we write $s_1 = s+\delta$ and $s_2 = s$. Considering $\hat{\beta}$ first,

$$\hat{\beta}(s,\delta) = \frac{s(s+\delta)[(s+\delta)\hat{L}(s+\delta)-s\hat{L}(s)]}{\delta+s(s+\delta)[\hat{L}(s+\delta)-\hat{L}(s)]}. \tag{5.33}$$

Dividing top and bottom by δ and allowing $\delta \rightarrow 0$, we obtain

$$\hat{\beta}(s) = \frac{s^2[s\hat{L}^{(1)}(s)+\hat{L}(s)]}{s^2\hat{L}^{(1)}(s)+1}. \tag{5.34}$$

For $\hat{\alpha}(s)$, we find from (5.32) that

$$e^{\hat{\alpha}(s,\delta)} = -\delta\frac{[(s+\delta)\hat{L}(s+\delta)-1][s\hat{L}(s)-1]}{\delta+s(s+\delta)(\hat{L}(s+\delta)-\hat{L}(s))}, \tag{5.35}$$

and applying the same method as before yields

$$e^{\hat{\alpha}(s)} = \frac{-[s\hat{L}(s)-1]^2}{1+s^2\hat{L}^{(1)}(s)}. \tag{5.36}$$

Note that this expression is valid only if $\hat{L}^{(1)}(s) < -s^2$.

These estimators are plotted in the following two figures. In each case the maximum likelihood estimator is indicated by a horizontal line. Overall, the transform estimators match the

Figure 5.5. Plot of $\alpha(s)$ Compared to the M.L.E..

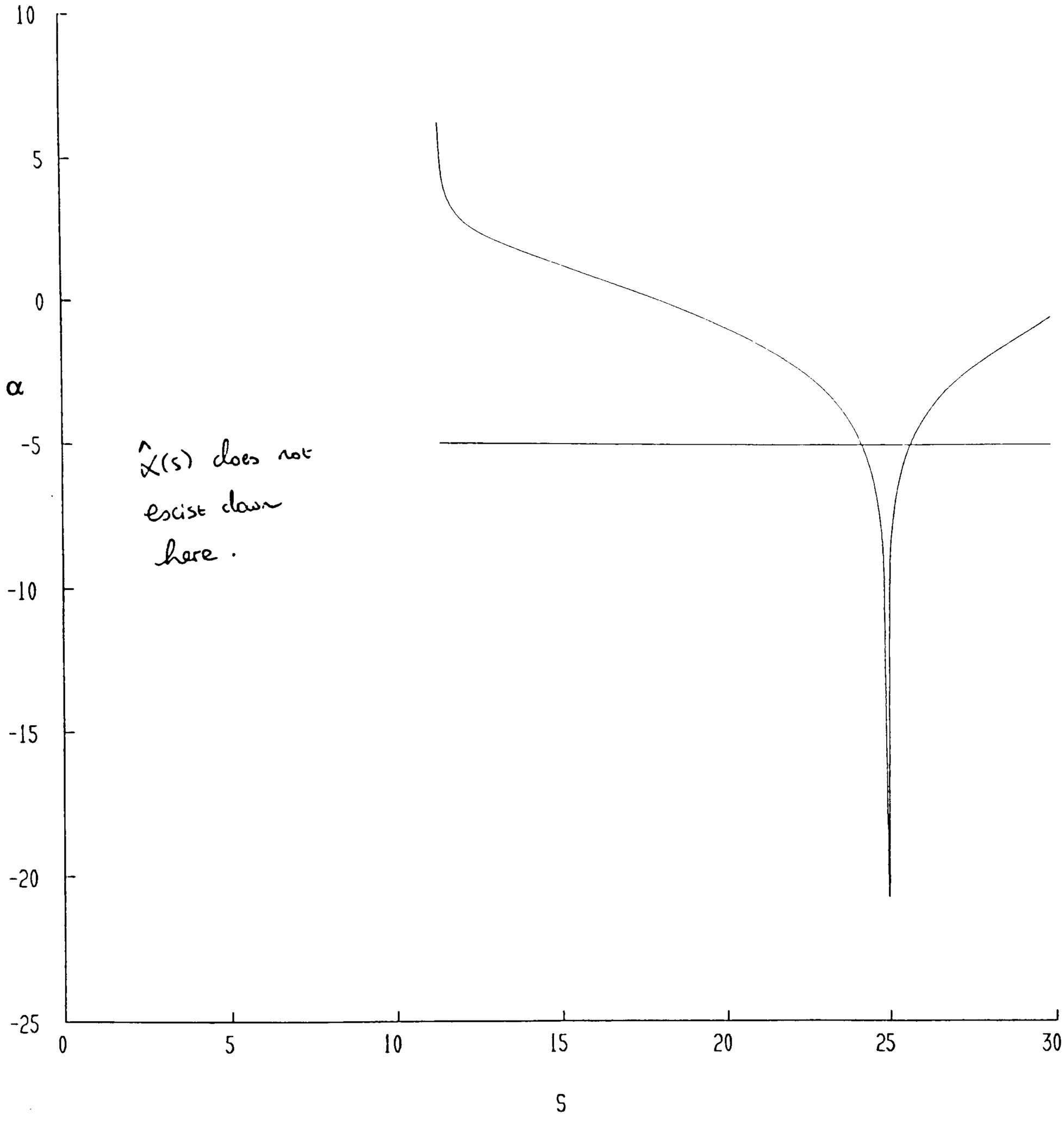
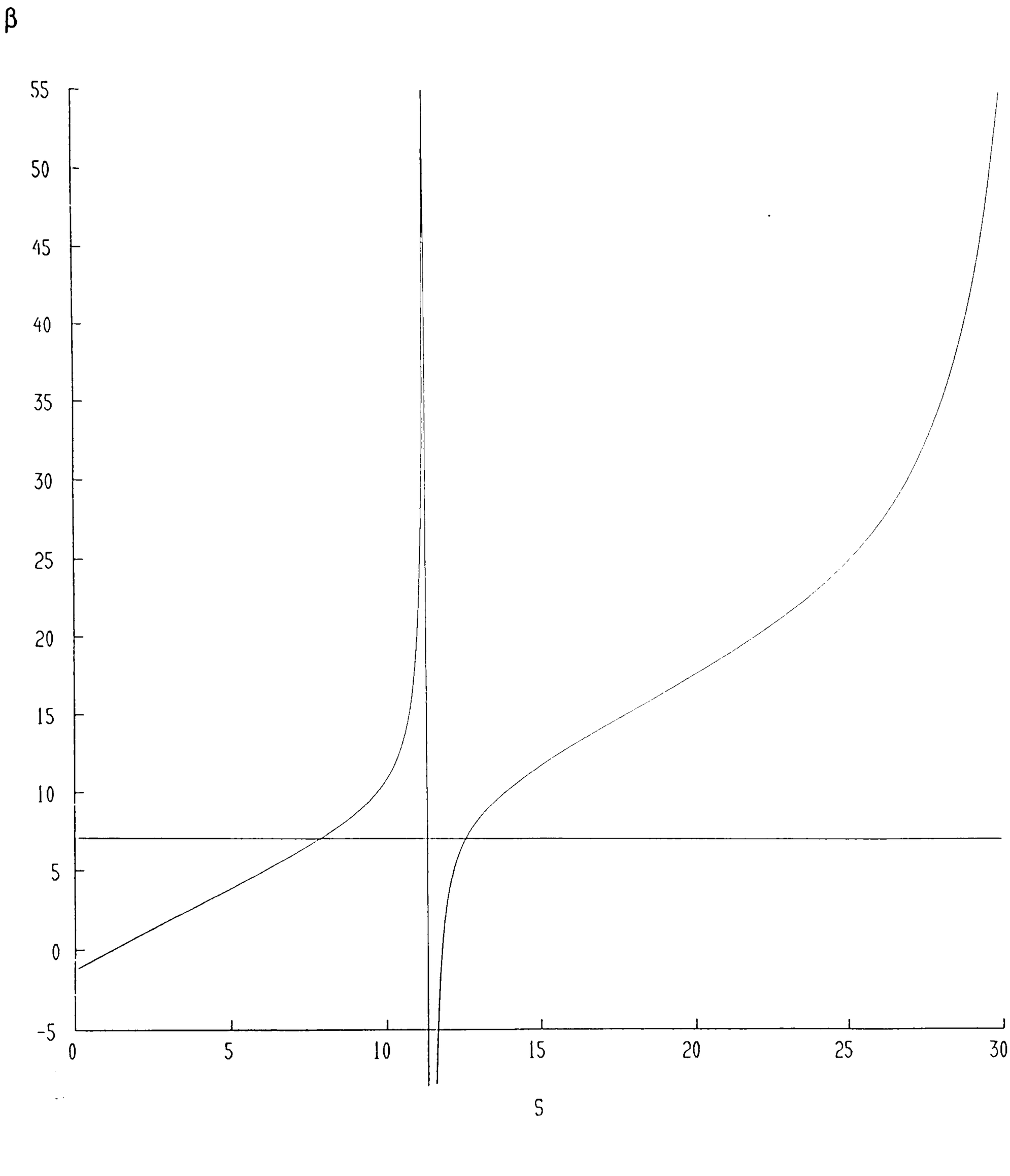


Figure 5.6. Plot of $\beta(s)$ Compared to the M.L.E..



m.l.e.s very poorly. When they do match, they do so in very different regions of s-space. It is clear that quadrature does not work well for this example; the precision of just 4 is inadequate to achieve a reasonable polynomial approximation to $f(x)$.

4.2. Riemann-Sum.

Parameter estimation stems from (5.29), which we re-quote here for ease of exposition.

$$\hat{L}(s)+ s^{-1}e^{-sy_n}[1+us] = s^{-1}+ e^{\alpha}(s-\beta)^{-1} ,$$

where $u = r_n(n_n - r_n)^{-1}$. The estimating equations may be written as

$$(s_i-\hat{\beta})(s_i\hat{L}_i+ e^{-sy_n}-1)+ s_i u e^{-s_i y_n} = s_i e^{\hat{\alpha}} \quad i=1,2. \tag{5.37}$$

We can simplify these a little by writing

$$w_{i1} = s_i\hat{L}_i+ e^{-s_i y_n}-1$$
$$w_{i2} = s_i u e^{-sy_n} ,$$

so that we obtain

$$s_i e^{\hat{\alpha}}+ w_{i1}\hat{\beta} = s_i w_{i1}+ w_{i2} \quad i=1,2. \tag{5.38}$$

In matrix terms, $\hat{\alpha}$ and $\hat{\beta}$ arise as the solutions to

$$\begin{bmatrix} s_1 & w_{11} \\ s_2 & w_{21} \end{bmatrix} \begin{bmatrix} e^{\hat{\alpha}} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} s_1 w_{11}+ w_{12} \\ s_2 w_{21}+ w_{22} \end{bmatrix} . \tag{5.39}$$

We therefore obtain the explicit estimators

$$\begin{bmatrix} e^{\hat{\alpha}} \\ \hat{\beta} \end{bmatrix} = (s_1 w_{21}- s_2 w_{11})^{-1} \begin{bmatrix} w_{11}w_{21}(s_1-s_2)+ w_{12}w_{21}- w_{11}w_{22} \\ s_1 s_2(w_{21}-w_{11})+ s_1 w_{22}- s_2 w_{12} \end{bmatrix} . \tag{5.40}$$

Referring back to Figure 5.4 we see that there is a match in the region of $s = 8.4$ between $\kappa(s)$ and $L_m(s)$. As such, we would expect to recover the m.l.e.s quite well in this region. Evaluating the estimators (5.40) reveals that there is a reasonable match with the m.l.e.s in the region $\underline{s} \in [8.2,8.4] \times [7.1,8.4]$. Outside this region $\hat{\alpha}(\underline{s})$ is particularly unstable. Indeed, it does not exist over large tracts of $\underline{s}$-space.

It seems sensible, then, to concentrate on this region. Our objective now is to compare the moment estimators derived above with the m.l.e.s. However, we require an objective criterion for

choosing $\underline{s}$ in order to do this. Since we have access to the log-likelihood $l(\alpha,\beta)$, we may form what is termed the *constrained* log-likelihood $l(\alpha(s),\beta(s))$. This term arises because the empirical and theoretical transforms are "constrained" to be equal. We have already obtained $\alpha(\underline{s})$ and $\beta(\underline{s})$ for $s_1 \neq s_2$, so it remains to consider the case $s_1 = s_2$. We employ the same technique used to find equations (5.34) and (5.36), although here we let $s_2 = s+\delta$ and $s_1 = s$. For $\hat{\beta}$ we obtain

$$\hat{\beta}(s,\delta) = \frac{s(s+\delta)\{s\delta^{-1}[\hat{L}(s+\delta)-\hat{L}(s)]+\hat{L}(s+\delta)+(u-1)\delta^{-1}[e^{-(s+\delta)}-e^{-sy_n}]\}}{s(s+\delta)\delta^{-1}[\hat{L}(s+\delta)-\hat{L}(s)]+\delta^{-1}[e^{-(s+\delta)y_n}-e^{-sy_n}]-e^{-sy_n}+1} . \quad (5.41)$$

Allowing $\delta \rightarrow 0$ we find

$$\hat{\beta}(s) = \frac{s^2[s\hat{L}^{(1)}(s)+\hat{L}(s)-y_n(u-1)e^{-sy_n}]}{1+s^2\hat{L}^{(1)}(s)-(1+y_n)e^{-sy_n}} . \quad (5.42)$$

Next, and rather more involved, for

$$e^{\hat{\alpha}(\underline{s})} = D^{-1}[(s_1\hat{L}_1+e^{-s_1y_n}-1)(s_2\hat{L}_2+e^{-s_2y_n}-1)(s_1-s_2)+ s_1ue^{-s_1y_n}(s_2\hat{L}_2+e^{-s_2y_n}-1)-s_2ue^{-s_2y_n}(s_1\hat{L}_1+e^{-s_1y_n}-1)] , \quad (5.43)$$

where $D = s_1w_{21}-s_2w_{11}$. It follows readily that

$$\lim_{\delta\rightarrow 0}\delta^{-1}D = 1+s^2\hat{L}^{(1)}(s)-(1+y_n)e^{-sy_n} . \quad (5.44)$$

The numerator of $e^{\hat{\alpha}(s,\delta)}$ is

$$[s\hat{L}(s)+e^{-sy_n}-1][(s+\delta)\hat{L}(s+\delta)+e^{-(s+\delta)y_n}-1](-\delta)+ sue^{-sy_n}[(s+\delta)\hat{L}(s+\delta)+e^{-(s+\delta)y_n}-1]-(s+\delta)ue^{-(s+\delta)y_n}[s\hat{L}(s)+e^{-sy_n}-1] . \quad (5.45)$$

Dividing by δ and taking $\lim_{\delta\rightarrow 0}$, we obtain

$$-[s\hat{L}(s)+e^{-sy_n}-1]^2-e^{-sy_n}u[s\hat{L}(s)+e^{-sy_n}-1]+ us\lim_{\delta\rightarrow 0}\delta^{-1}[a(s)b(s+\delta)-a(s+\delta)b(s)] , \quad (5.46)$$

where $a(s) = e^{-sy_n}$ and $b(s) = s\hat{L}(s)+e^{-sy_n}-1$. Both $a(s)$ and $b(s)$ are infinitely differentiable, so we may write

$$a(s+\delta) = a(s)+\delta a'(s)+O(\delta^2)$$
$$b(s+\delta) = b(s)+\delta b'(s)+O(\delta^2) , \quad (5.47)$$

by Taylor-series expansion. The limit term of (5.46) may be re-expressed, using (5.47), as

$$\lim_{\delta\to 0}\delta^{-1}\{\,a(s)[b(s)+\delta b'(s)+O(\delta^2)]-b(s)[a(s)+\delta a'(s)+O(\delta^2)]\,\}$$
$$=a(s)b'(s)-b(s)a'(s)\,. \tag{5.48}$$

Thus,

$$e^{\hat{\alpha}(s)}=\frac{use^{-sy_n}[s\hat{L}^{(1)}(s)+(1+sy_n)\hat{L}(s)-y_n]-[s\hat{L}(s)+e^{-sy_n}-1]^2-ue^{-sy_n}[s\hat{L}(s)+e^{-sy_n}-1]}{1+s^2\hat{L}^{(1)}(s)-e^{-sy_n}(1+y_n)}\,. \tag{5.49}$$

A simple grid-search of the constrained log-likelihood revealed a maximum at $\underline{s}=(8.38,7.50)'$, where $L(\alpha(s),\beta(s))=-119.856$ corresponding to the maximum log-likelihood quoted in Table 5.2. We therefore recover the m.l.e.s by this procedure. For the translated data we have

$$\hat{\alpha}=-2.716265\quad\hat{\beta}=7.071732\,, \tag{5.50}$$

so for the original data, correct to 3sf,

$$\hat{\alpha}=-4.84\quad\hat{\beta}=7.07\,. \tag{5.51}$$

The small sample properties of these estimators were investigated, via a simulation study, by Laurence et al. (1986). These are of particular importance because such samples are frequently encountered in practice. The transform variable was selected by maximizing the constrained likelihood, and it was found that the transform approach can out-perform maximum likelihood in these circumstances.

We have used the Quantal Assay model to illustrate transform procedures in the stochastic modelling context, and do not dwell here on measures of error for parameter estimators. This will, however, be an important topic in the more detailed work to follow. In the discussion so far we have witnessed the poor performance of quadrature when the sampling points are not designed. In this case the Riemann-Sum approach to transform construction is clearly superior. We move on now to consider the One-Compartment Open model, a particular example of the models discussed in section 5.2.2. In this context we observe the benefits of conducting designed experiments. A particular difference between the work so far and that to follow concerns selection of the

transform variable, as we shall see.

5. The One-Compartment Open Model.

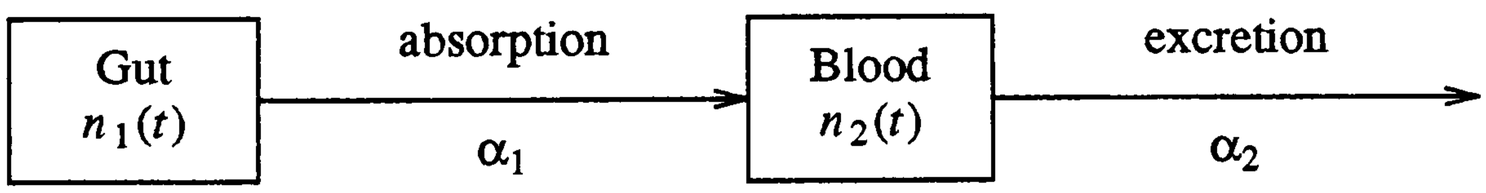
Compartment models describe the movement of material between particular states known as compartments. These could represent, for example, grades of employment or stages of a chemical reaction. It is assumed that the rates of flow between compartments follow first order kinetics, so that the rate of transfer to a receiving, or *sink,* compartment is proportional to the concentration in the supplying, or *source* compartment. A collection of inter-connected compartments is referred to as a *system* (of compartments).

These models have found a wide variety of uses, including chemical kinetics, manpower planning and pharmacokinetics. The material represented might be the molecules of some chemical, individual workers and drug concentration respectively. Having discussed compartment models in general terms, we consider a specific

Example 5.2.

Wagner's (1967) experiment, as reviewed by Shah (1976), is often quoted in the compartment modelling literature. A Tetracycline compound was given orally to a subject and its subsequent concentration in serum was measured over a period of 16 hours. The following deterministic compartment model was fitted to the data:

Figure 5.7. Compartment Model Fitted to Wagner (1967) Data.



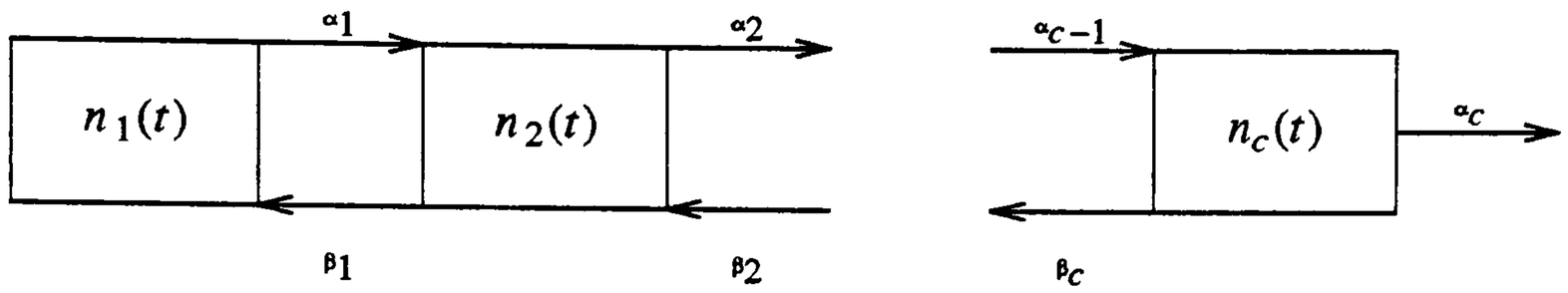
Denoting $\underline{n}(t) = (n_1(t), n_2(t))'$, then we have the initial condition $\underline{n}(t) = (\alpha_3, 0)'$, so that α_3 represents the initial dose of material, in general, introduced into the system. In this deterministic scenario, the movement of material through the system is governed by the system of differential equations

$$\frac{dn_1(t)}{dt} = -\alpha_1 n_1(t)$$

$$\frac{dn_2(t)}{dt} = \alpha_1 n_1(t) - \alpha_2 n_2(t)\,. \tag{5.52}$$

The above example is a special, important case of the general class of open-compartment models. These may be represented in general as

Figure 5.8. General Open System.



The distinguishing feature of this type of model is that it allows for material to exit from a single end compartment. In the employment model this end compartment could represent retirement from the work force for example.

In the work to follow we assume that the compartments contain discrete elements, or may reasonably be approximated as such. We go on now to consider the absorption-excretion model of Example 5.2; the One-Compartment Open (OCO) model. This model has found a number of important applications, particularly in the investigation of chemical and drug kinetics. See, for example, Macheras et al. (1992). Applications of compartment models in this sort of work are not limited to open systems however, as demonstrated by Wald et al. (1991).

We have thus far discussed only the deterministic version of the OCO model. A more realistic model might be one that includes a stochastic element. If we allow transitions from compartments to be Markov processes, with transition intensities given by the αs, then the system (5.52) describes the behaviour of the expected compartment contents. Thus, if the contents are denoted by the random vector $\underline{N}(t) = (N_1(t), N_2(t))'$, then we have $E(N_i(t)) = n_i(t)$ $i=1,2$. Solving equations (5.52), we obtain

$$\begin{aligned} n_1(t) &= \alpha_3 e^{-\alpha_1 t} \\ n_2(t) &= \alpha_1 \alpha_3 [\alpha_1 - \alpha_2]^{-1} \left[e^{-\alpha_2 t} - e^{-\alpha_1 t}\right]. \end{aligned} \tag{5.53}$$

The case $\alpha_1 > \alpha_2$ is of most practical interest in this thesis. A stochastic treatment will be given in sub-section 6.2 of this work on the OCO model.

Typically, as in Example 5.1, interest will focus on $N_2(t)$, so we assume that data $\{N_2(t_1), \ldots, N_2(t_n)\}$ are collected in order to estimate the parameter vector $\underline{\alpha} = (\alpha_1, \alpha_2, \alpha_3)'$. To this end, we may model the observations from the second compartment as

$$N_2(t) = n_2(t) + \varepsilon(t)\,, \tag{5.54}$$

in the form of section 5.2.2. As such, all our earlier work on transform construction and parameter estimation applies. We therefore propose to base estimation on the Laplace transform

$$\begin{aligned} L(s;\underline{\alpha}) &= \int_0^\infty e^{-st}\, n_2(t)dt \\ &= \frac{\alpha_1\alpha_3}{(\alpha_1+s)(\alpha_2+s)}\,. \end{aligned} \tag{5.55}$$

5.1. Parameter Estimation.

There are a number of distinct approaches to consider here. We may choose multiple values of s to obtain enough estimating equations, or we could alternatively employ just one s-value. In addition, we deal with the cases where the initial dose (α_3) may or may not be be known. In the work to follow we first consider estimation employing as many values of s as unknown parameters, subsequently obtaining the estimators based on a single value of s by the limiting procedures employed earlier.

5.1.1. Initial Dose Known.

Our task here is to estimate the transition intensities α_1 and α_2. Choosing two values of s, s_1 and s_2, we have from (5.55) that

$$(\alpha_1+s_i)(\alpha_2+s_i) = \alpha_1\alpha_3/L(s_i) \quad i=1,2. \tag{5.56}$$

For ease of exposition we do not, for the moment, work with empirical quantities, but substitute these at a later stage. Now, expanding equations (5.56) gives

$$\alpha_1\alpha_2 + s_i(\alpha_1+\alpha_2) + s_i^2 = \alpha_1\alpha_3/L_i \quad i=1,2,$$

in our usual notation. Thus,

$$(s_1 - s_2)(\alpha_1 + \alpha_2) + s_1^2 - s_2^2 = \alpha_1 \alpha_3 (L_1^{-1} - L_2^{-1})$$
$$\Rightarrow \quad \alpha_1 + \alpha_2 + s_1 + s_2 = \alpha_1 g(\underline{s}),$$

where

$$g(\underline{s}) = \alpha_3 (L_1^{-1} - L_2^{-1})(s_1 - s_2)^{-1}.$$
$$\therefore \quad \alpha_1 (1 - g(\underline{s}) + \alpha_2 + s_1 + s_2) = 0$$
$$\Rightarrow \quad \hat{\alpha}_2 = (\hat{g}(\underline{s}) - 1)\hat{\alpha}_1 - s_1 - s_2 . \tag{5.57}$$

Substituting for α_2 in the first equation of system (5.56) we obtain the quadratic equation

$$(\hat{g} - 1)\hat{\alpha}_1^{\,2} - \omega\,\hat{\alpha}_1 - s_1 s_2 = 0 , \tag{5.58}$$

where $\omega = s_1 + s_2 + \alpha_3 / L_1 - s_1 \hat{g}$. Thus,

$$\hat{\alpha}_1 = 0.5(\hat{g} - 1)^{-1} [\omega \pm (\omega^2 + 4(\hat{g}(\underline{s}) - 1) s_1 s_2)^{1/2}] . \tag{5.59}$$

We choose the root which is in the proper range such that both estimators are positive.

Having obtained the estimators based on two values of s, we now allow $\underline{s} \rightarrow s\underline{1}$, where $\underline{1}$ represents a vector of 1s. The only limit of real interest is

$$\lim_{s_1 \to s_2} \hat{g}(\underline{s}) = \lim_{s_1 \to s_2} \alpha_3 (s_1 - s_2)^{-1} (\hat{L}_1^{-1} - \hat{L}_2^{-1}) . \tag{5.60}$$

It is convenient to re-cast this as

$$\lim_{\delta \to 0} \hat{g}(s + \delta, s) = \lim_{\delta \to 0} \alpha_3 \delta^{-1} [\hat{L}(s + \delta)^{-1} - \hat{L}(s)^{-1}]$$

$$= \alpha_3 \frac{d}{ds} [\hat{L}(s)^{-1}]$$

$$= - \alpha_3 \hat{L}'(s) / \hat{L}(s)^2 . \tag{5.61}$$

Making the necessary substitutions in (5.59) and (5.57) yields the forms of $\hat{\alpha}_1(s)$ and $\hat{\alpha}_2(s)$ respectively. We move on now to the case where α_3 is unknown.

5.1.2. Initial Dose Unknown.

Since there is now an extra parameter to estimate, we require three values s_1, s_2, s_3 of s. Let

$$\kappa_i = L(s_i)/L(s_{i+1}) = \frac{(s_{i+1} + \alpha_1)(s_{i+i} + \alpha_2)}{(s_i + \alpha_1)(s_i + \alpha_2)} \qquad i = 1, 2. \tag{5.62}$$

Again, we neglect empirical quantities for the moment and form the equations

$$\alpha_1\alpha_2(1-\kappa_i)+(\alpha_1+\alpha_2)(s_{i+1}-\kappa_i s_i)=\kappa_i s_i^2-s_{i+1}^2 \qquad i=1,2.$$

In matrix terms,

$$\begin{bmatrix} s_2-\kappa_1 s_1 & 1-\kappa_1 \\ s_3-\kappa_2 s_2 & 1-\kappa_2 \end{bmatrix} \begin{bmatrix} \alpha_1+\alpha_2 \\ \alpha_1\alpha_2 \end{bmatrix} = \begin{bmatrix} \kappa_1 s_1^2-s_2^2 \\ \kappa_2 s_2^2-s_3^2 \end{bmatrix}.$$

Substituting empirical quantities then, we obtain

$$\begin{bmatrix} \hat{\alpha}_1+\hat{\alpha}_2 \\ \hat{\alpha}_1\hat{\alpha}_2 \end{bmatrix} = [(1-\hat{\kappa}_1)(s_3-s_2\hat{\kappa}_2)-(1-\hat{\kappa}_2)(s_2-s_1\hat{\kappa}_1)]^{-1}\ \underline{u}\ , \tag{5.63}$$

where

$$\underline{u}=\begin{bmatrix} (1-\hat{\kappa}_1)(\hat{\kappa}_2 s_2^2-s_3^2)-(1-\hat{\kappa}_2)(\hat{\kappa}_1 s_1^2-s_2^2) \\ (s_3-s_2\hat{\kappa}_2)(\hat{\kappa}_1 s_1^2-s_2^2)-(s_2-s_1\hat{\kappa}_1)(\hat{\kappa}_2 s_2^2-s_3^2) \end{bmatrix}.$$

We obtain an estimator of α_3 from

$$\hat{\alpha}_3\hat{\alpha}_1(s_3+\hat{\alpha}_1)^{-1}(s_3+\hat{\alpha}_2)^{-1}=\hat{L}_3\ . \tag{5.64}$$

To find $\hat{\alpha}_1$ and $\hat{\alpha}_2$ explicitly, note that we may form the quadratic equation

$$z^2-(\hat{\alpha}_1+\hat{\alpha}_2)z+\hat{\alpha}_1\hat{\alpha}_2=0 \tag{5.65}$$

$$\Rightarrow\ (z-\hat{\alpha}_1)(z-\hat{\alpha}_2)=0$$

$$\Rightarrow\ \hat{\alpha}_1,\hat{\alpha}_2=\frac{\hat{\alpha}_1+\hat{\alpha}_2}{2}\pm[(\frac{\hat{\alpha}_1+\hat{\alpha}_2}{2})^2-\hat{\alpha}_1\hat{\alpha}_2]^{1/2}\ , \tag{5.66}$$

so that we can find $\hat{\alpha}_1$ and $\hat{\alpha}_2$ given their sum and product, choosing $\hat{\alpha}_1>\hat{\alpha}_2$.

We may employ similar techniques as before to investigate the behaviour of these estimators as $\underline{s}\rightarrow s\underline{1}$, but we omit the detail here. The case α_3 known seems to be the more usual, since the initial dose in, for example, a scientific experiment will be known unless there is some loss due to an accident. In the light of this, we concentrate attention, although not exclusively, on the details of the first case.

Before proceeding further, we discuss an alternative approach to estimating the transition intensities when the initial dose is known.

5.1.3. A Modified Transform Approach.

We use here a similar modified transform to that of Chapter 4, and base estimation on

$$\psi(s) = -\frac{d}{ds}\log sL(s;\underline{\alpha}) = (s+\alpha_1)^{-1}+(s+\alpha_2)^{-1}, \tag{5.67}$$

so this transform possesses a particularly simple form. Given two distinct values of s we obtain estimating equations based on the method-of-moments as

$$(s_i\hat{\psi}_i-1)(\hat{\alpha}_1+\hat{\alpha}_2)+\hat{\psi}_i\hat{\alpha}_1\hat{\alpha}_2 = s_i(2-\hat{\psi}_i s_i) \quad i=1,2. \tag{5.68}$$

Solving these equations we obtain

$$\begin{bmatrix}\hat{\alpha}_1+\hat{\alpha}_2\\ \hat{\alpha}_1\hat{\alpha}_2\end{bmatrix} = [\hat{\psi}_2(s_1\hat{\psi}_1-1)-\hat{\psi}_1(s_2\hat{\psi}_2-1)]^{-1}\underline{u}, \tag{5.69}$$

where

$$\underline{u} = \begin{bmatrix} s_1\hat{\psi}_2(2-\hat{\psi}_1 s_1)-s_2\hat{\psi}_1(2-\hat{\psi}_2 s_2) \\ s_2(2-\hat{\psi}_2 s_2)(s_1\hat{\psi}_1-1)-s_1(2-\hat{\psi}_1 s_1)(s_2\hat{\psi}_2-1)\end{bmatrix}.$$

We may, therefore, form the quadratic equation

$$(z-\hat{\alpha}_1)(z-\hat{\alpha}_2)=0$$

in the same way as before; see (5.65).

Quadrature was employed in transform construction, by methods to be discussed in detail in the next section. The estimators were subsequently evaluated over an $\underline{s}$-grid, revealing that they are very unstable. This seems to be due to ill-conditioned estimating equations, as the following discussion suggests.

The empirical and theoretical transforms match quite well for $s\in(3.0,5.0)$, so we focus attention on this region. The table below clearly shows this matching, for values of s close to an intersection of ψ and $\hat{\psi}$, for the case $\alpha_1 = 0.10$ and $\alpha_2 = 0.05$.

Table 5.5. Match of $\hat{\psi}$ to ψ.

i	s_i	$\hat{\psi}_i$	ψ_i	$1-\hat{\psi}_i/\psi_i$
1	3.78	0.5185	0.5188	$7.21\text{x } 10^{-4}$
2	3.80	0.5171	0.5162	$-1.75\text{x } 10^{-3}$

Using these s-values.,

$$\begin{bmatrix} 0.95993 & 0.5185 \\ 0.96498 & 0.5171 \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1+\hat{\alpha}_2 \\ \hat{\alpha}_1\hat{\alpha}_2 \end{bmatrix} = \begin{bmatrix} 0.1514646 \\ 0.1330760 \end{bmatrix},$$

$$\Rightarrow \quad \begin{bmatrix} \hat{\alpha}_1+\hat{\alpha}_2 \\ \hat{\alpha}_1\hat{\alpha}_2 \end{bmatrix} = \begin{bmatrix} -2.353 \\ 4.648 \end{bmatrix}$$

which is clearly ridiculous.

Since finding $\hat{\alpha}_1$ and $\hat{\alpha}_2$ from their sum and product reduces to finding the intersection of two lines, the angle between these lines, θ say, is an important quantity. It turns-out that $\theta = 0.19^o$ in this case, so the problem seems to be ill-conditioned. For $s_1 = 3.0$ and $s_2 = 5.0$, at the boundary of the matching region, and using values of ψ rather than $\hat{\psi}$, we find $\theta = 12.3^o$. There still seems to be some risk of ill-conditioning here.

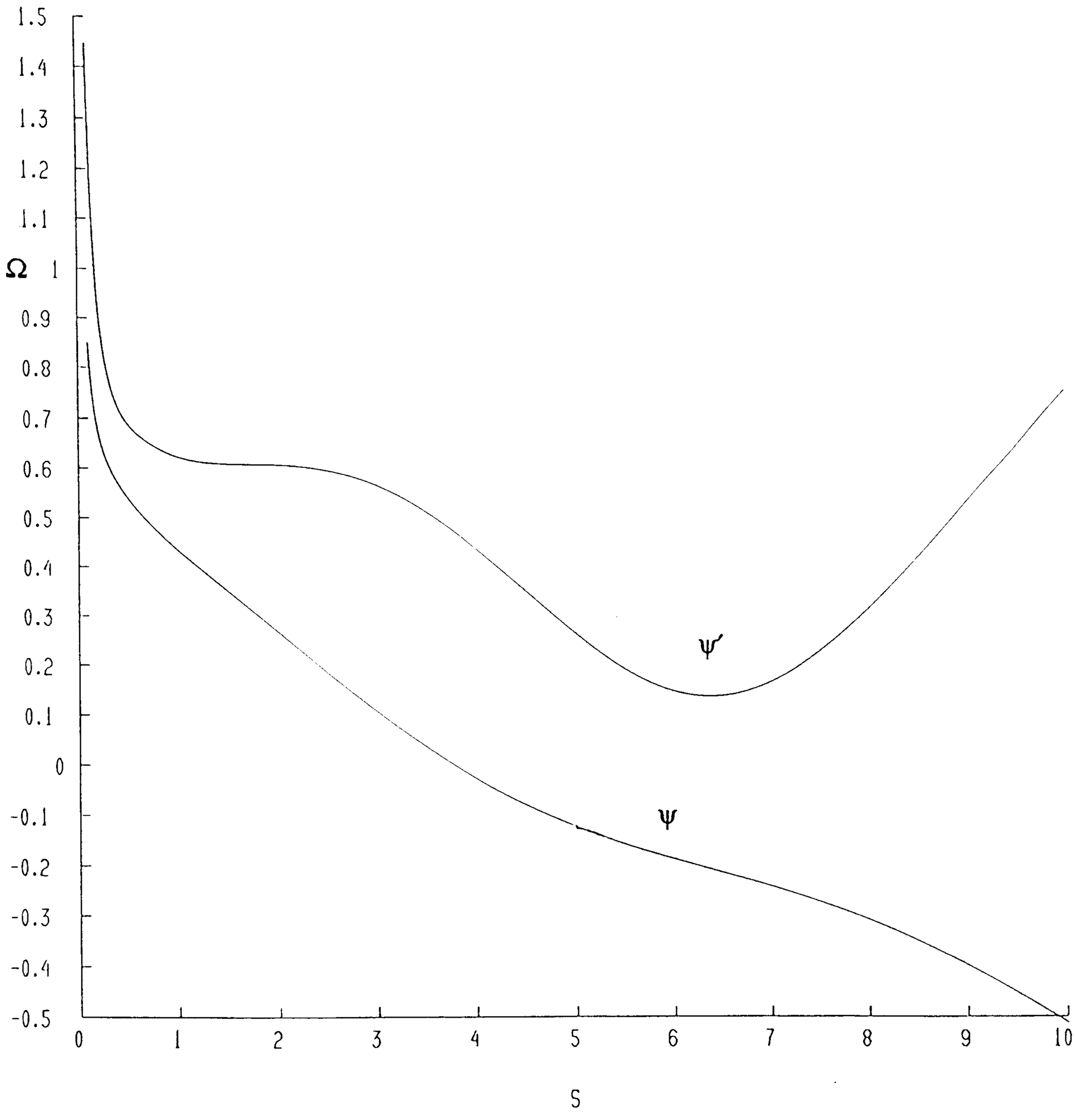
The difficulties with this approach suggest that we might concentrate on the intersection of the empirical and theoretical transforms. Utilizing one value of s, we require a match between both the transforms and first derivatives. To investigate these quantities, we define

$$\Omega(s) = 1 - \frac{\textit{Empirical Transform}}{\textit{Theoretical Transform}},$$

which is depicted below.

We look for regions where Ω is close to 0, but we also require these to be coincident. This is clearly not achieved here; it seems that $\hat{\psi}$ is a very unstable transform. This result is very much in accord with earlier work in this thesis, and seems to be a general lesson of the approach based on modified transforms. Modified transforms often yield, by design, explicit estimators. However, we shouldn't expect to receive this benefit without penalty, as has been shown in a number of

Figure 5.9. Ω for ψ and ψ′.



examples discussed in this thesis.

5.2. Selecting the Transform Variable.

We seek here some relatively simple objective criterion for selecting the transform variable, $\underline{s}$. Many possibilities exist, but probably the most straightforward is Least-Squares. We therefore choose $\underline{s}$, denoting the least-squares function LS, as

$$LS(\underline{s}_{opt}) = \min_{\underline{s}} \sum_{j=1}^{n} [N_2(t_j) - \hat{N}_2(t_j)]^2$$

$$= \min_{\underline{s}} \sum_{j=1}^{n} [N_2(t_j) - n_2(t_j; \underline{\alpha} = \underline{\alpha}(\underline{s}))]^2 . \qquad (5.70)$$

This will, in general, involve us in a p-dimensional search if we have p parameters to estimate. However, the transform approach holds out the possibility of reducing this dimension. Indeed, by allowing $\underline{s} \rightarrow s\underline{1}$ we need only conduct a one-dimensional search. It will clearly be important to establish whether we can exploit this property whilst producing efficient estimators.

It is desirable to make comparisons with a standard technique to act as a benchmark, and the natural choice here is Ordinary Least-Squares, viz:

$$LS(\hat{\underline{\alpha}}_{ols}) = \min_{\underline{\alpha}} \sum_{j=1}^{n} [N_2(t_j) - n_2(t_j; \underline{\alpha})]^2 . \qquad (5.71)$$

We now have in place techniques of parameter estimation and transform variable selection. Before we can investigate the performance of transform methods, it will be useful to be able to simulate data from this model.

5.3. Simulating the One-Compartment Open Model.

There are two distinct flows of material present in this model. The first is entry of material into the open compartment, which we model as a Poisson process of rate $\alpha_1 N_1(t)$. The second flow is out of the system, modelled as a Poisson process of rate $\alpha_2 N_2(t)$. We simulate the overall process as follows:

Step 1: Simulate an event from a Poisson process having rate parameter $\alpha_1 N_1(t) + \alpha_2 N_2(t)$.

Step 2: We assign this event to one of the two possibilities noted above, with probabilities given below.

Table 5.6. Assignment Probabilities for Simulation Routine.

Event	Probability
Exit from Source	$\dfrac{\alpha_1 N_1(t)}{\alpha_1 N_1(t) + \alpha_2 N_2(t)}$
Exit from Sink	$\dfrac{\alpha_2 N_2(t)}{\alpha_1 N_1(t) + \alpha_2 N_2(t)}$

The process begins with the initial conditions $N_1(0) = \alpha_3$ and $N_2(0) = 0$. A flowchart to implement this simulation routine follows in Figure 5.10 below, with notes detailing the NAG routines employed.

Three examples of output from this routine follow in Figures 5.11-13 below. The initial dose was of one hundred elements, so $\alpha_3 = 100$, with transition intensities $\alpha_1 = 0.2$, $\alpha_2 = 0.1$. The solid line indicates the expected contents, $n_2(t; \underline{\alpha})$. It is clear from these results that a wide variety of patterns are possible from the same parameter combination, suggesting that we require a relatively robust estimation procedure. Indeed, Frome & Yakatan (1980) emphasize the need for robustness of model-fitting in this context. It is comforting to note that the time to peak concentration is not consistently over- or under-estimated.

There is one further refinement required. The results of the above simulation correspond to a process observed continuously, whereas in practice we only observe the process at the sampling points $\underline{t}$. We may clearly obtain the relevant data, in a straightforward fashion, directly from the continuously observed process, although we do not dwell on the details. To conclude this section, we note that the techniques employed here to simulate the OCO model may clearly be extended to more complicated situations.

Figure 5.10. Simulation Algorithm for the One-Compartment Open Model.

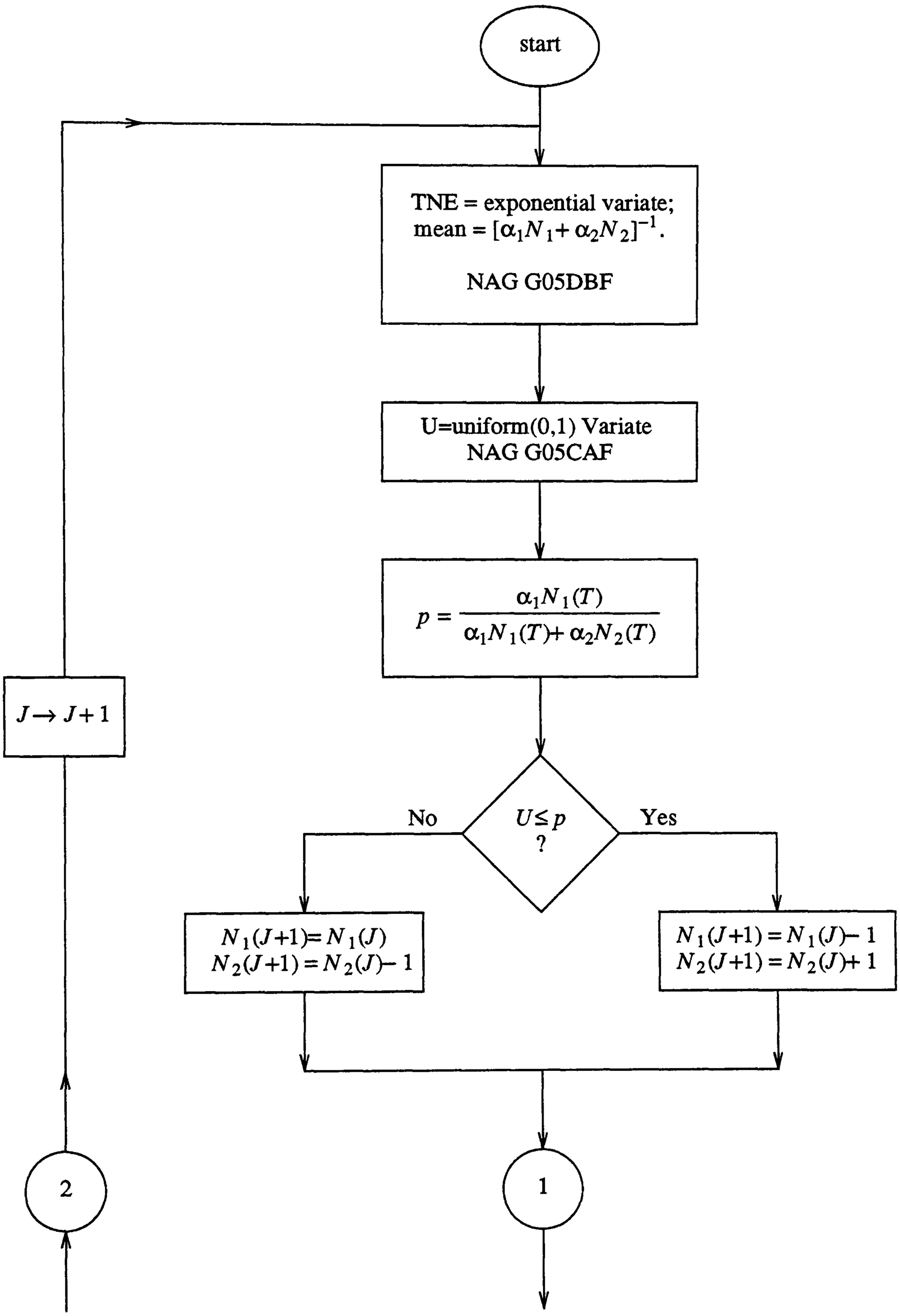
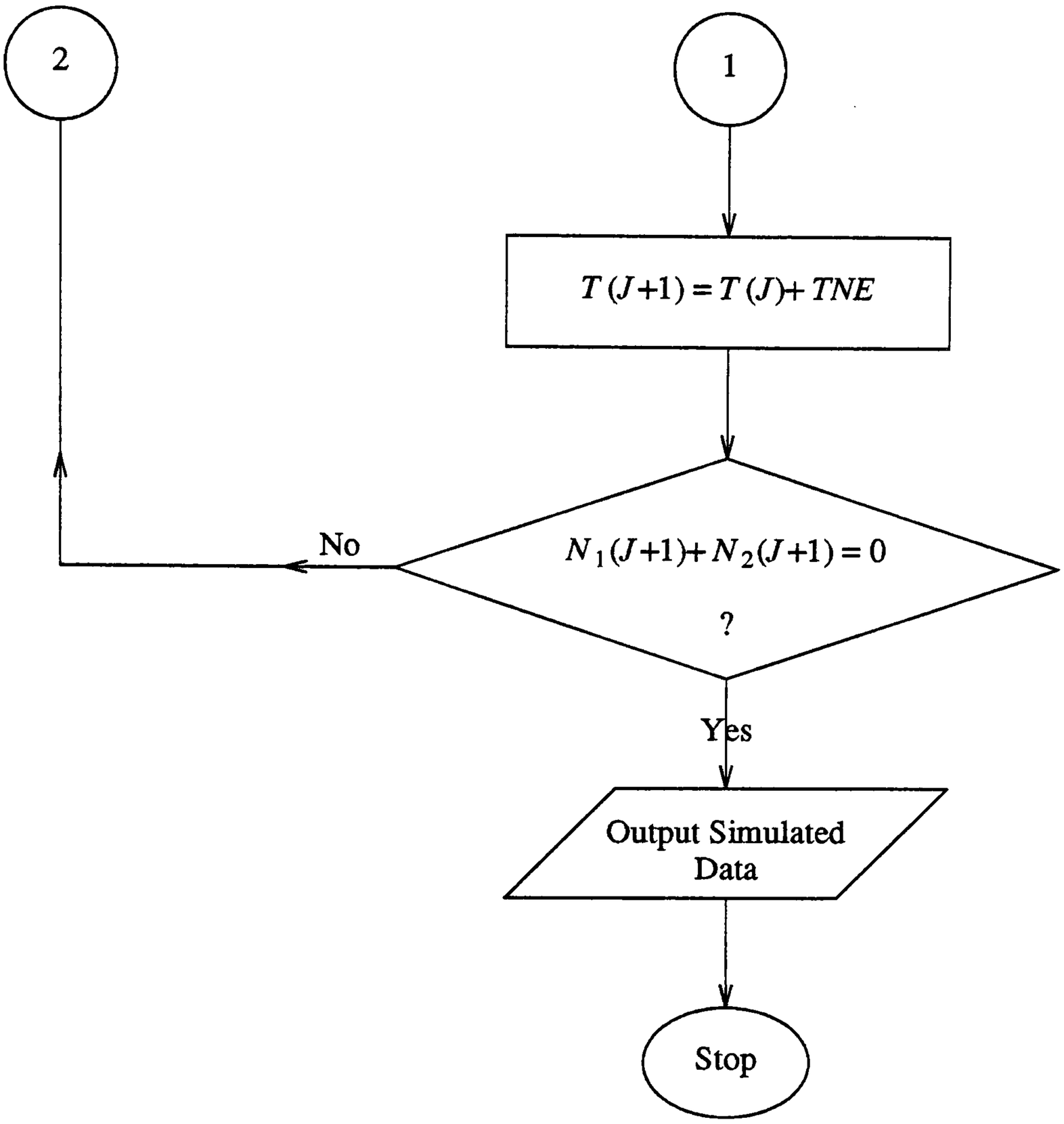


Figure 5.10. - Continued.



Variables Used: T - Time Index, updated when each transition takes place; TNE - Inter-event time; U - Uniform variate used for deciding which type of transition has taken place; J - Event counter.

Figure 5.11. Simulated Data from the OCO Model : 1.

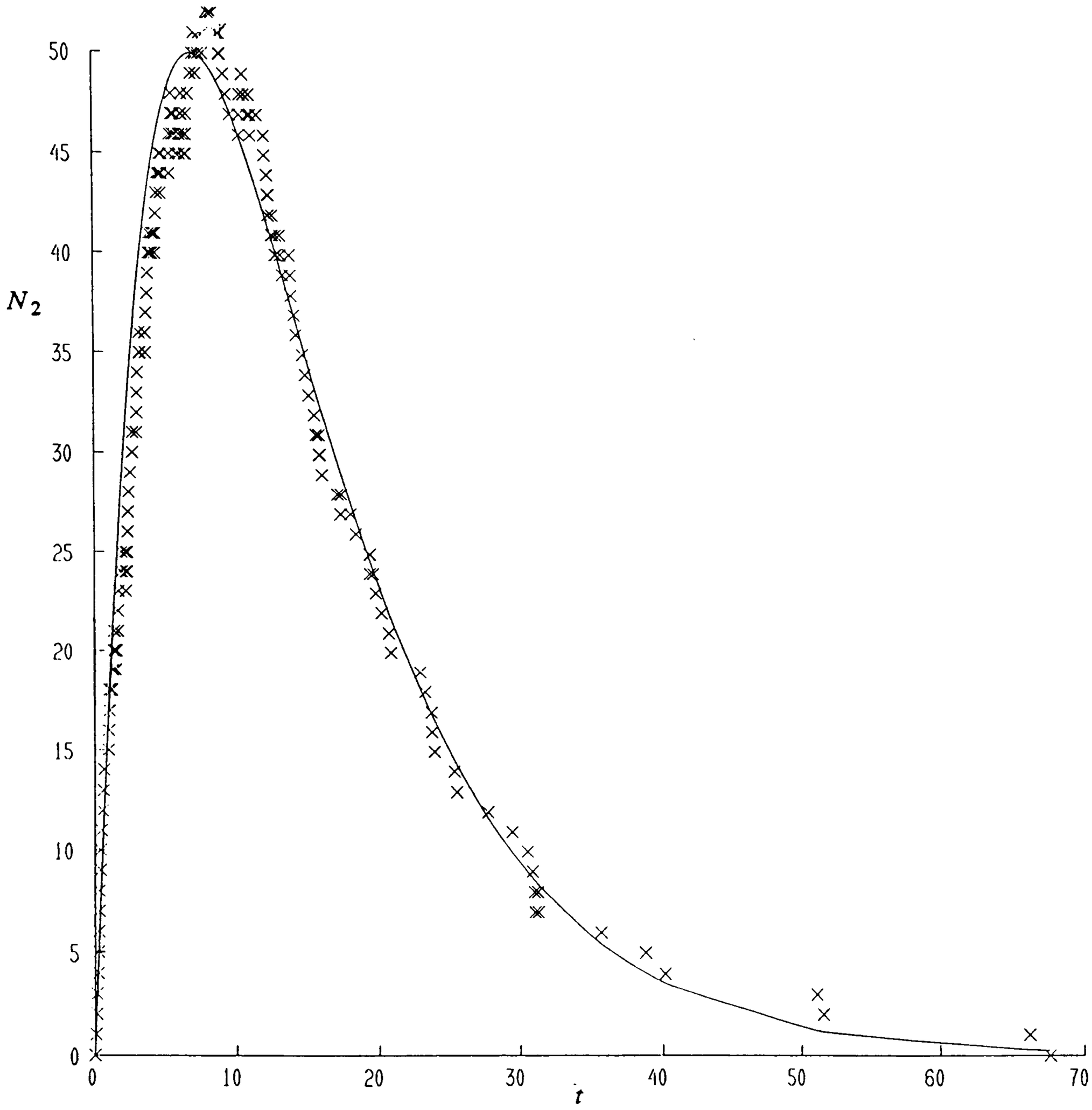


Figure 5.12. Simulated Data from the OCO Model : 2.

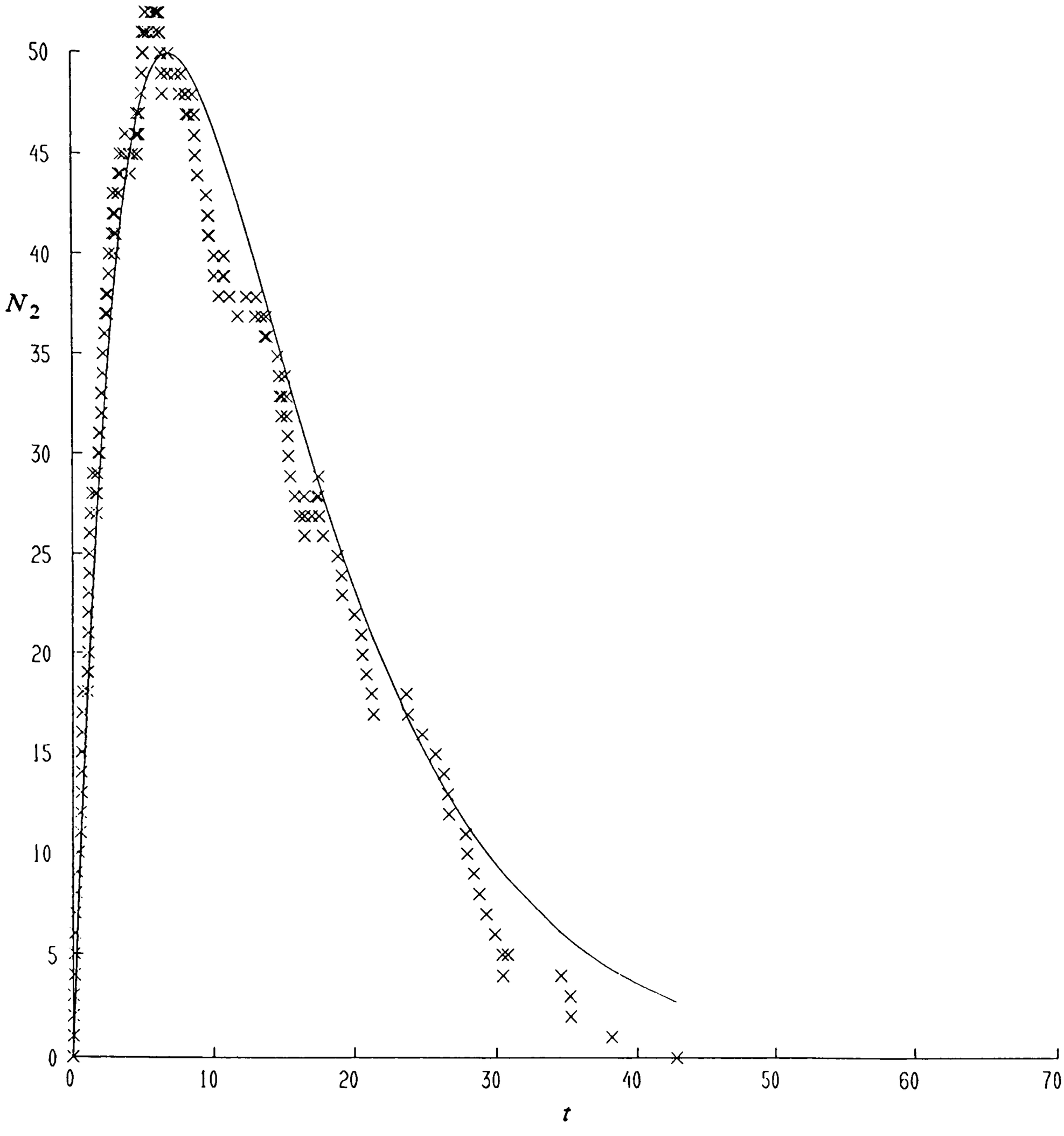
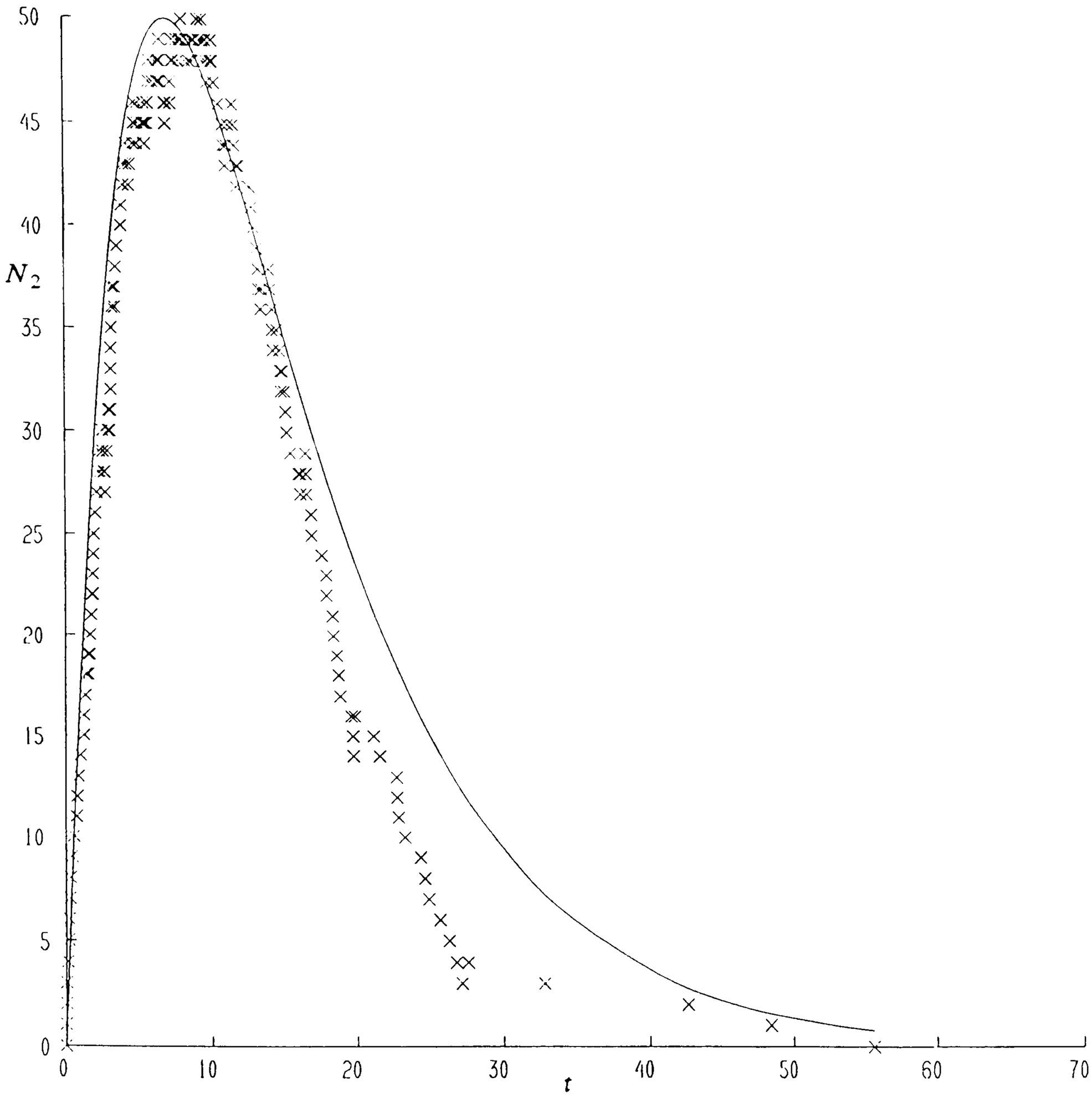


Figure 5.13. Simulated Data from the OCO Model : 3.



5.4. Transform Construction and Performance.

Recalling the definition leading to equation (5.55), the transform employed here is

$$L(s;\underline{\alpha}) = \int_0^\infty e^{-st}\, n_2(t;\underline{\alpha})dt.$$

In the spirit of our earlier work, this may be written as

$$L(s;\underline{\alpha}) = \int_0^\infty e^{-t}\, f(t;\underline{\alpha})dt\ , \tag{5.72}$$

where $f(t;\underline{\alpha}) = e^{(1-s)t} n_2(t;\underline{\alpha})$. Both methods of transform construction will then yield an ELT of the form

$$\hat{L}(s) = \sum_{k=1}^{n} h_k(s) N_2(t_k)\ . \tag{5.73}$$

The precise form of $h_k(s)$ depends on the method of construction, so we look at the detail for each method next.

5.4.1. Quadrature

We are here in the desirable position of being able to plan our experimental design in advance of data collection. With this in mind, we may apply Gauss-Laguerre quadrature to (5.72), yielding weights $\underline{w}$ and, critically, the sampling points $\underline{t}$. We obtain a scheme of precision $2n-1$, with empirical transform

$$\hat{L}(s) = \sum_{k=1}^{n} w_k e^{(1-s)t_k} N_2(t_k)\ . \tag{5.74}$$

A visual inspection of Gauss-Laguerre weights shows that they all lie between 0 and 1, in stark contrast to the undesigned case of section 5.3.1.1. There does seem to be considerable scope for optimism that this method might perform very much better. To support this view we note that $f(t)$ is a mixture of exponentials here, a quantity which we might reasonably hope to approximate well by a finite polynomial.

We now move on to an initial investigation of the quality of the ELT (5.74). To this end, we define the ratio

$$\psi(s) = 1 - \hat{L}(s)/L(s;\underline{\alpha}), \tag{5.75}$$

which is plotted in Figures 5.14-18 below for a variety of sample sizes and particular sets of simulated data. NAG routine D01BBF was employed to obtain the quadrature weights, which restricts calculation of $\underline{w}$ to specified sample sizes. As such, we select a broad range of sample sizes from those available, choosing $n \in \{5,12,16,20,24\}$ in this thesis. Numerical difficulties (underflow) were experienced when attempting to obtain $\underline{w}$ for larger values of n, although this is not a great limitation here since a sample of size 24 is quite large for this application.

For the higher transition intensities studied there exists the possibility that all material will have exited the system before sampling is complete, especially for the larger sampling schemes. As a result, the figures below are not complete for all possible cases. We instead plot the smallest, $n = 5$, case along with the largest complete sample. If possible, an intermediate case is plotted as well.

Apart from sample size, we must also choose the values of $\underline{\alpha}$. To this end, α_3 was fixed at 100 elements, and the transition intensities were varied over "small", "medium" and "large" values, constrained so that $\alpha_1 > \alpha_2$. Thus, we investigate $\alpha_1 \in \{0.10,0.50,0.95\}$ and $\alpha_2 \in \{0.05,0.45,0.90\}$, giving six cases in total. It was found that no sampling scheme was complete for the case $\alpha_1 = 0.95$, $\alpha_2 = 0.90$; the material leaves the system too quickly. This is a problem which we return to in the more detailed work to come.

Turning now to the results of plotting $\psi(s)$, note that we are looking for regions of close agreement between the ELT and its theoretical counterpart, $L(s;\underline{\alpha})$, indicated by $\psi(s) \approx 0$. In the neighbourhood of such an s, we might hope to recover the true parameter values quite well, so these regions are particularly important. ψ is plotted over quite a wide region, but it is clear that the best agreement occurs for quite small values of s, although clearly away from the origin. The $n = 5$ case performs almost uniformly badly, the exception being for $\alpha_1 = 0.50$, $\alpha_2 = 0.05$ where the ELT lies within 5% of L over quite a large region. An interesting feature is that global performance does not necessarily improve with sample size. Overall, these results seem quite encouraging. It seems that we obtain an estimator which performs quite well for a reasonable range of s-

values.

Figure 5.14. $\psi(s)$ for $\alpha_1 = 0.10$, $\alpha_2 = 0.05$.

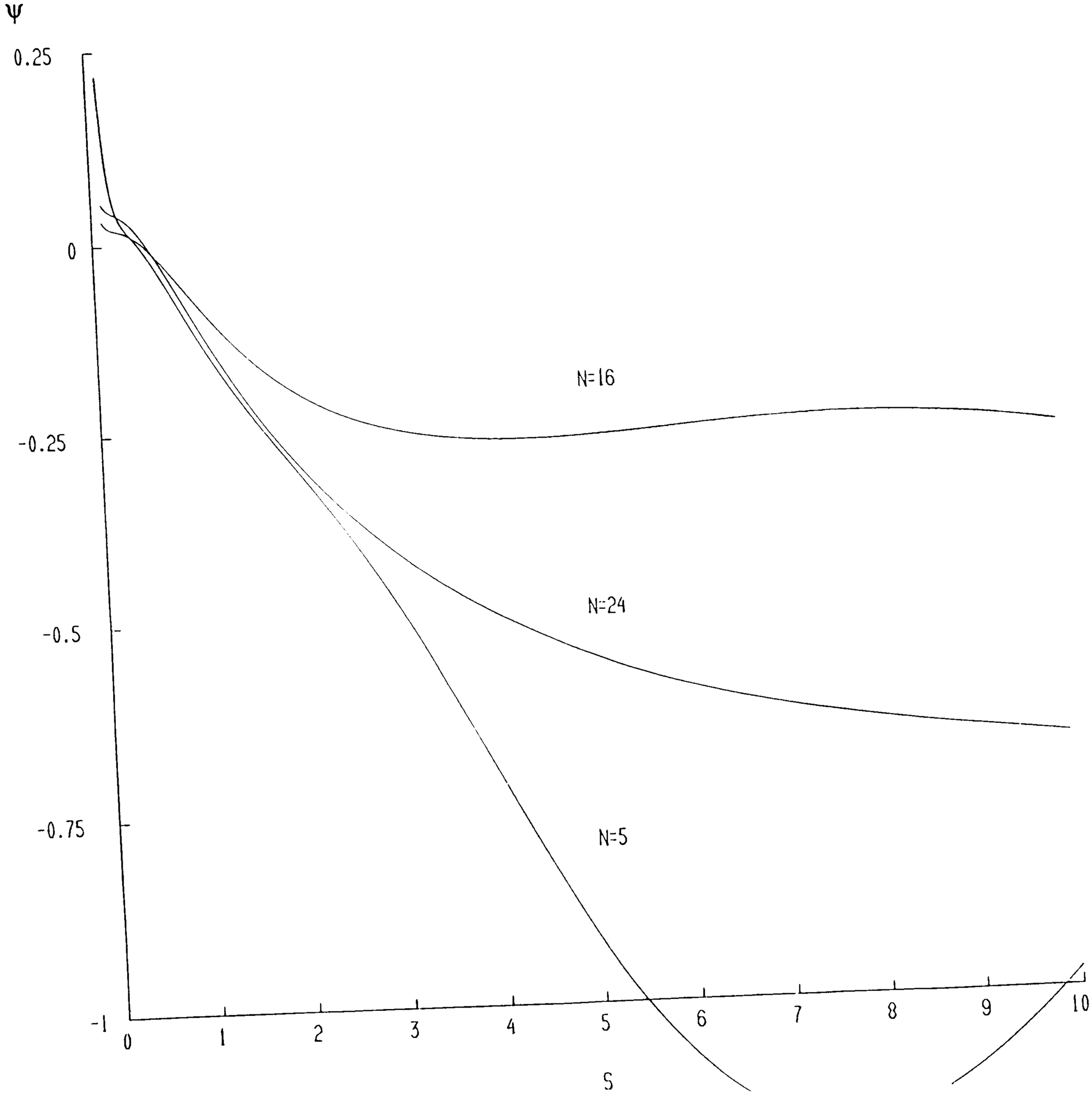


Figure 5.15. $\psi(s)$ for $\alpha_1 = 0.50,\ \alpha_2 = 0.05$.

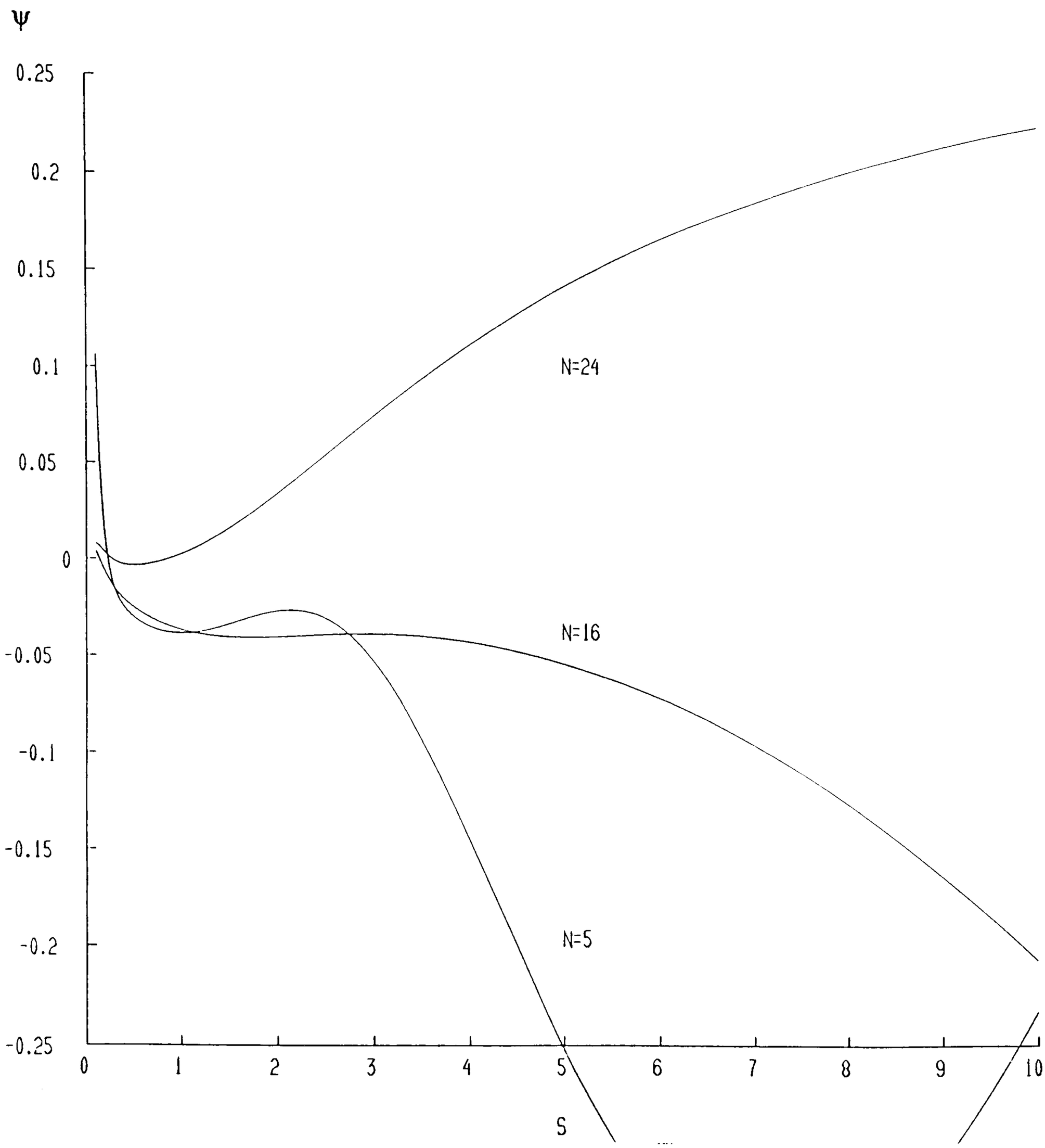


Figure 5.16. $\psi(s)$ for $\alpha_1 = 0.50$, $\alpha_2 = 0.45$.

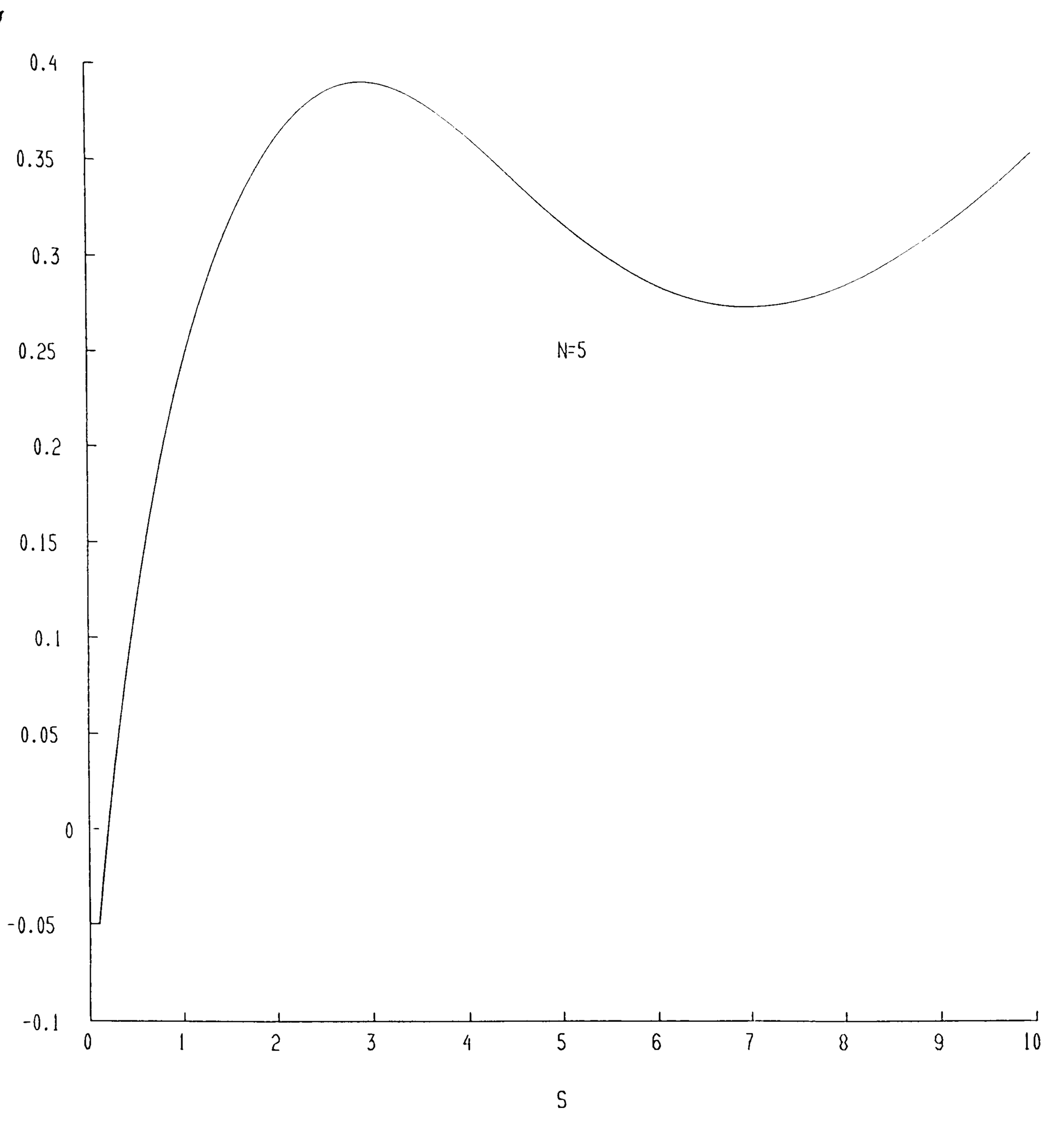


Figure 5.17. $\psi(s)$ for $\alpha_1 = 0.95$, $\alpha_2 = 0.05$.

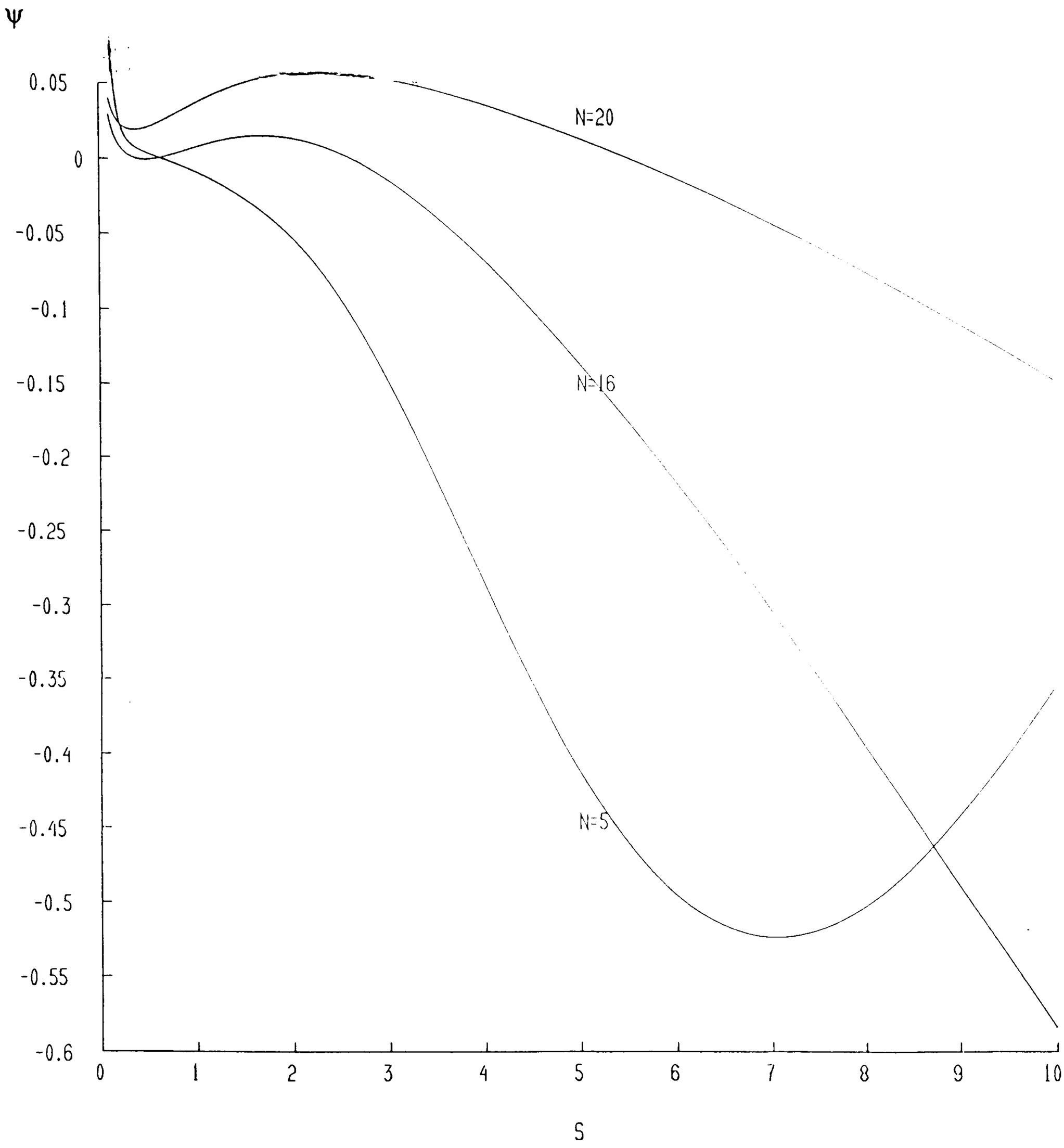
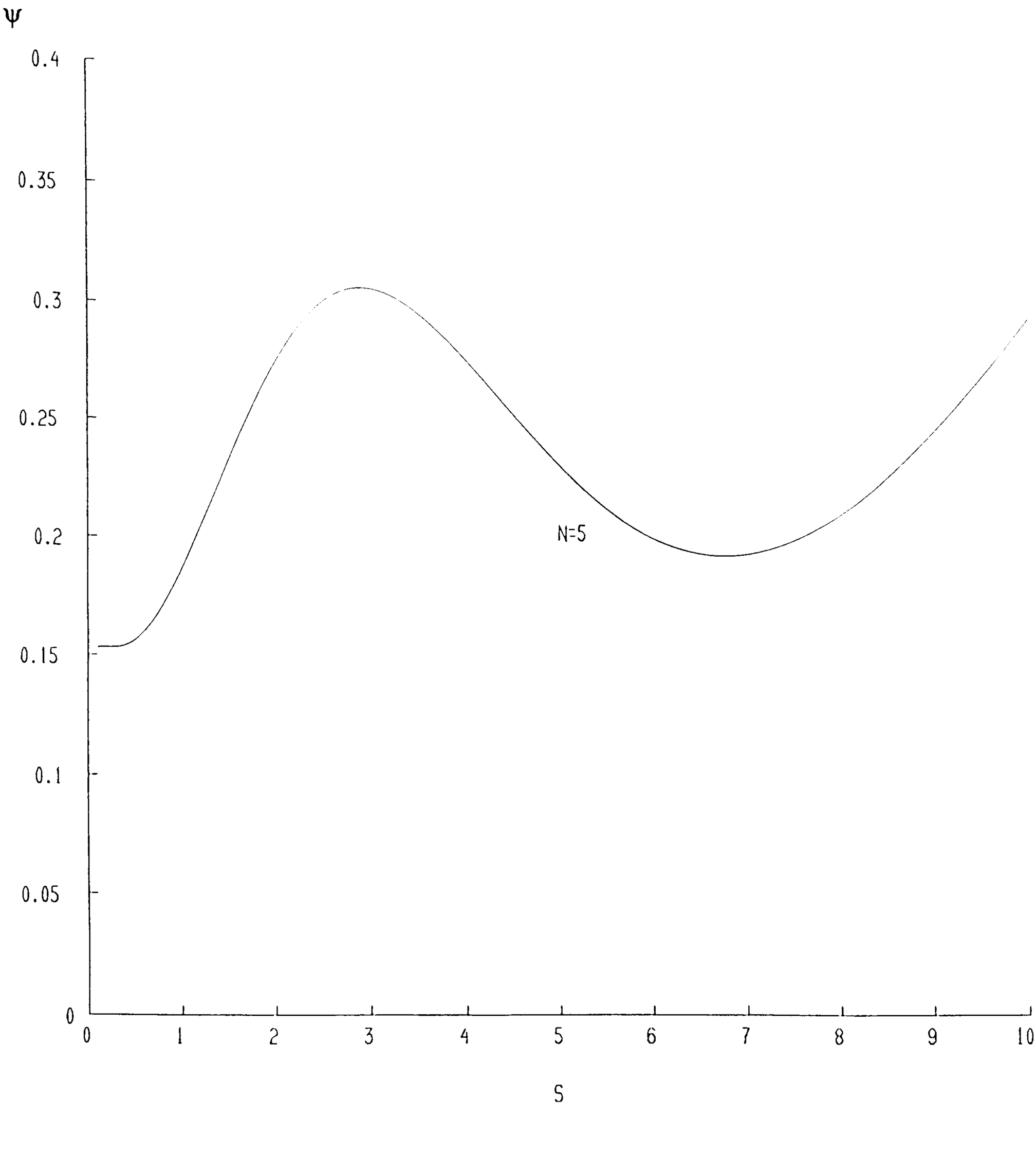


Figure 5.18. $\psi(s)$ for $\alpha_1 = 0.95$, $\alpha_2 = 0.45$.



We proceed now to consider our alternative method of transform construction.

5.4.2. Riemann-Sum.

Recalling (5.25) yields the form of $h_k(s)$ in (5.73) for this method of transform construction. We form the ELT as

$$\hat{L}(s) = s^{-1} \sum_{k=1}^{n} [e^{-sc_{i-1}} - e^{-sc_i}] N_2(t_k) , \tag{5.76}$$

where

$$c_0 = 0,\ c_n = \infty \text{ and } c_j = 0.5[t_j + t_{j+1}] \quad 1 \le j \le n-1 .$$

This method of construction does not aid us, as did quadrature, in matters of experimental design. To provide a sensible comparison between the competing methods, then, we employ the Gauss-Laguerre sampling points.

Our earlier work on the Quantal Assay model suggests that a corrected version of (5.76) might be superior. Since we are using Gauss-Laguerre sampling points, t_1 will tend to be quite close to the origin. This avoids the need for lower end-correction. However, this still leaves the requirement for an upper end-correction to the ELT. To do this, we set $c_n = N_2(t_n)$ so that $\hat{L}(s)$ estimates the incomplete Laplace transform

$$\int_0^{N(t_n)} e^{-st}\, n_2(t; \underline{\alpha})dt. \tag{5.77}$$

In order to estimate $L(s; \underline{\alpha})$, we need also to evaluate $I(s; \underline{\alpha})$ in the notation of section 5.3.2.1. Parameter estimation would then be based on an equation analogous to (5.28), which works out to

$$\hat{L}(s) + \frac{\alpha_1 \alpha_3}{(s+\alpha_1)(s+\alpha_2)} \{ (\alpha_1 - \alpha_2)^{-1} [(s+\alpha_1) e^{-(s+\alpha_2) N_2(t_n)} - (s+\alpha_2) e^{-(s+\alpha_1) N_2(t_n)} - 1] \} = 0. \tag{5.78}$$

In the Quantal Assay example we were able to introduce an approximation which considerably simplified matters, but this does not seem to be possible here. As such, we use the ELT (5.76), although we should be wary when t_n is small.

Our initial investigation of $\hat{L}(s)$ is exactly the same as that for quadrature. To this end, $\psi(s)$ is plotted for the same parameter combinations and sample sizes in Figures 5.19-23 below. We compare and contrast the performance of the Riemann-sum estimator versus quadrature in case-by-case fashion next.

(i) $\alpha_1 = 0.10,\ \alpha_2 = 0.05.$

Unsurprisingly the $n = 24$ scheme seems to provide the best performance, with both methods performing similarly for this case. Of particular interest is the $n = 16$ case, where quadrature seems to do much better. Both methods perform badly for the smallest sample size, although quadrature seems to perform better for small values of s.

(ii) $\alpha_1 = 0.50,\ \alpha_2 = 0.05.$

Riemann-sum does extremely well for the larger sample sizes, although quadrature compares very favourably towards smaller values of s.

(iii) $\alpha_1 = 0.50,\ \alpha_2 = 0.45.$

Both techniques perform very badly, since we could only obtain a complete sample for the smallest sample size. In order to investigate this combination of transition intensities we would require more material in the system. That is, a larger value of α_3; we return to this question at a later stage.

(iv) $\alpha_1 = 0.95,\ \alpha_2 = 0.05.$

Taking into account the different scales, there seems to be little to choose between the methods, at least for the larger samples. For the $n = 5$ case the Riemann-sum approach is clearly inferior.

(v) $\alpha_1 = 0.95,\ \alpha_2 = 0.45.$

This a case where we cannot obtain sufficient information to yield a good estimate of $L(s; \underline{\alpha})$, due again to insufficient material in the system.

We are now in a position to examine the performance of the parameter estimators themselves. As well as statistical performance, we are also interested in the practical aspects of fitting the OCO model. This motivates the need for the following section.

Figure 5.19. $\psi(s)$ for $\alpha_1 = 0.10$, $\alpha_2 = 0.05$.

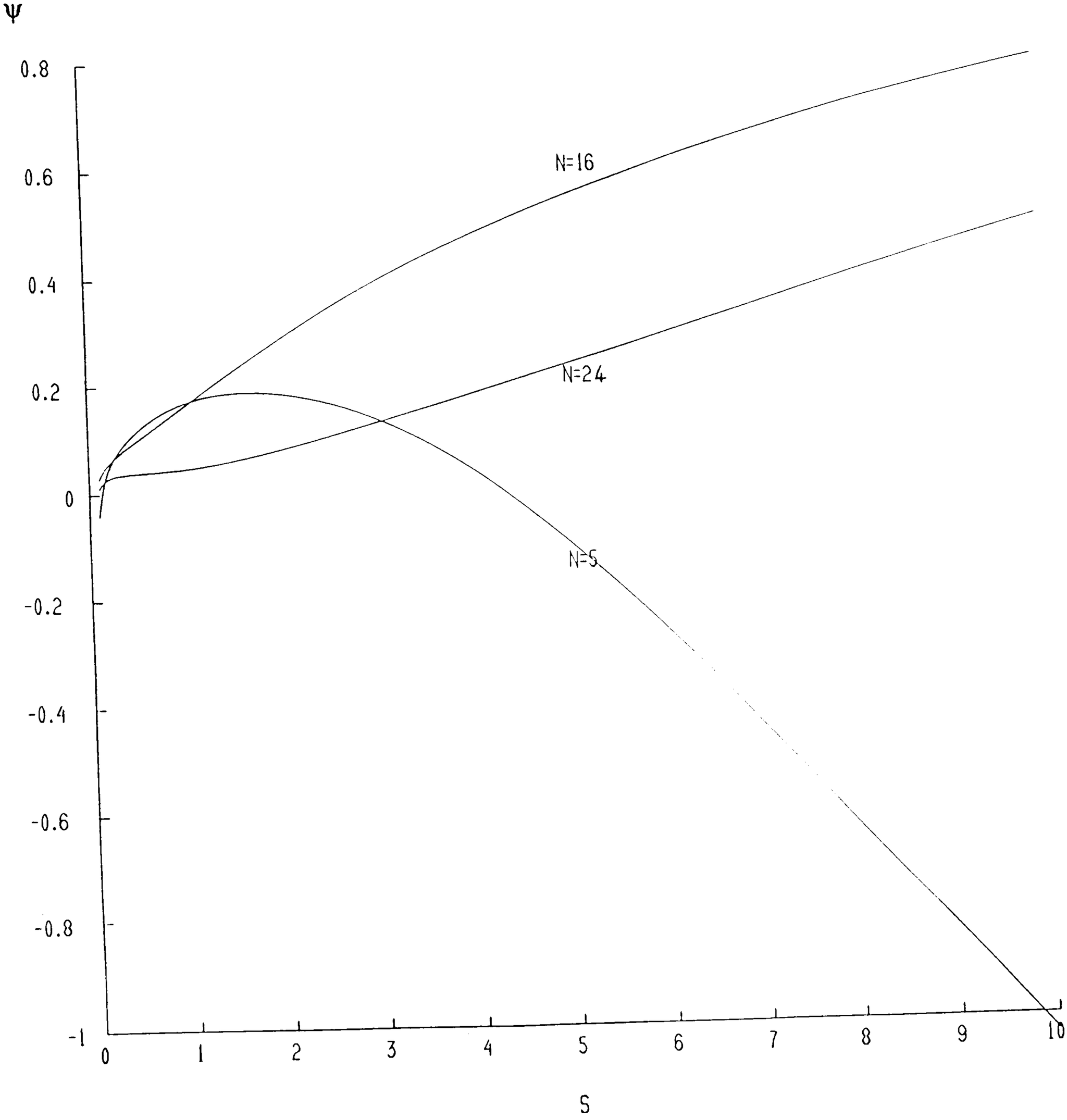


Figure 5.20. $\psi(s)$ for $\alpha_1 = 0.50$, $\alpha_2 = 0.05$.

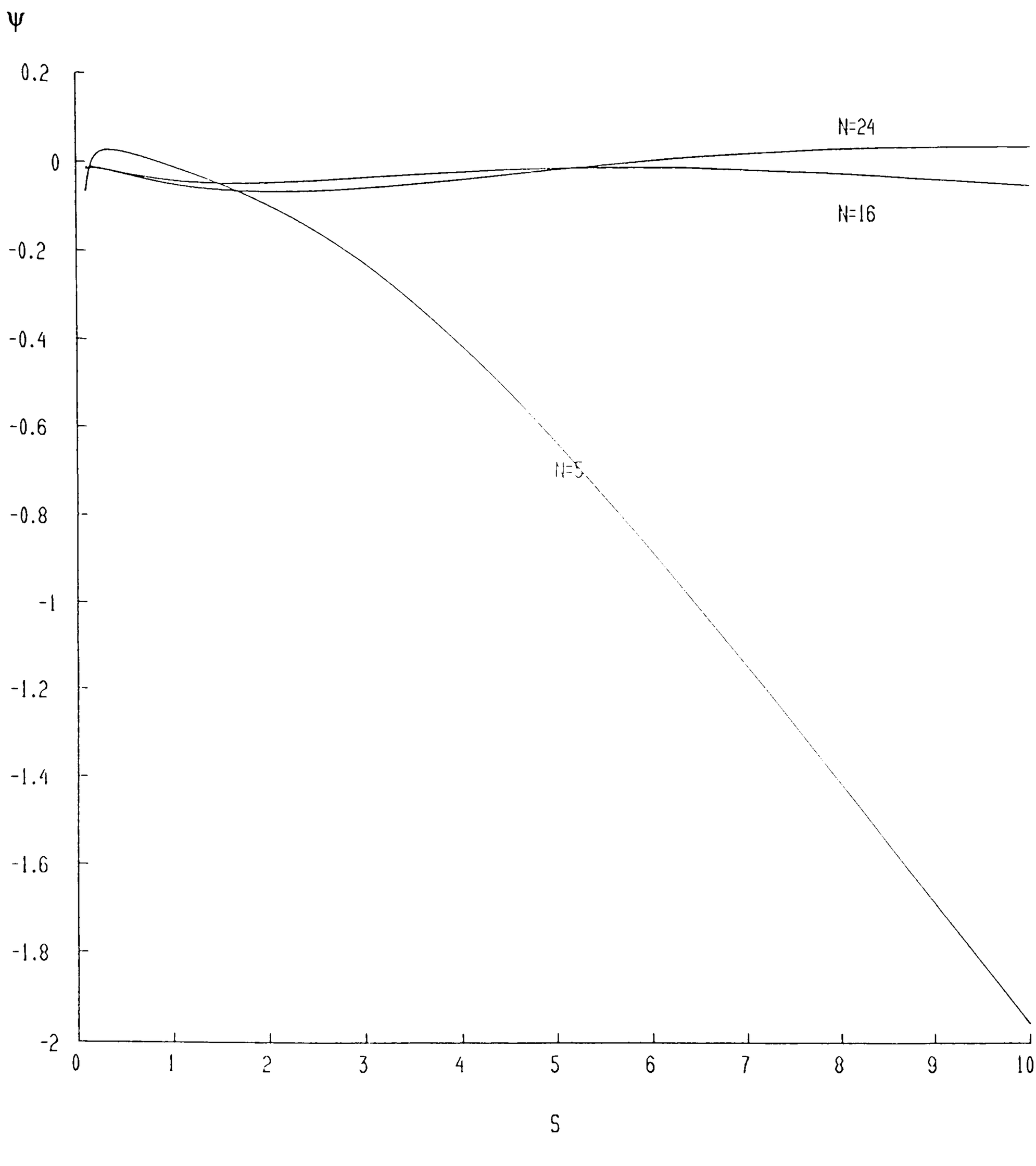


Figure 5.21. $\psi(s)$ for $\alpha_1 = 0.50$, $\alpha_2 = 0.45$.

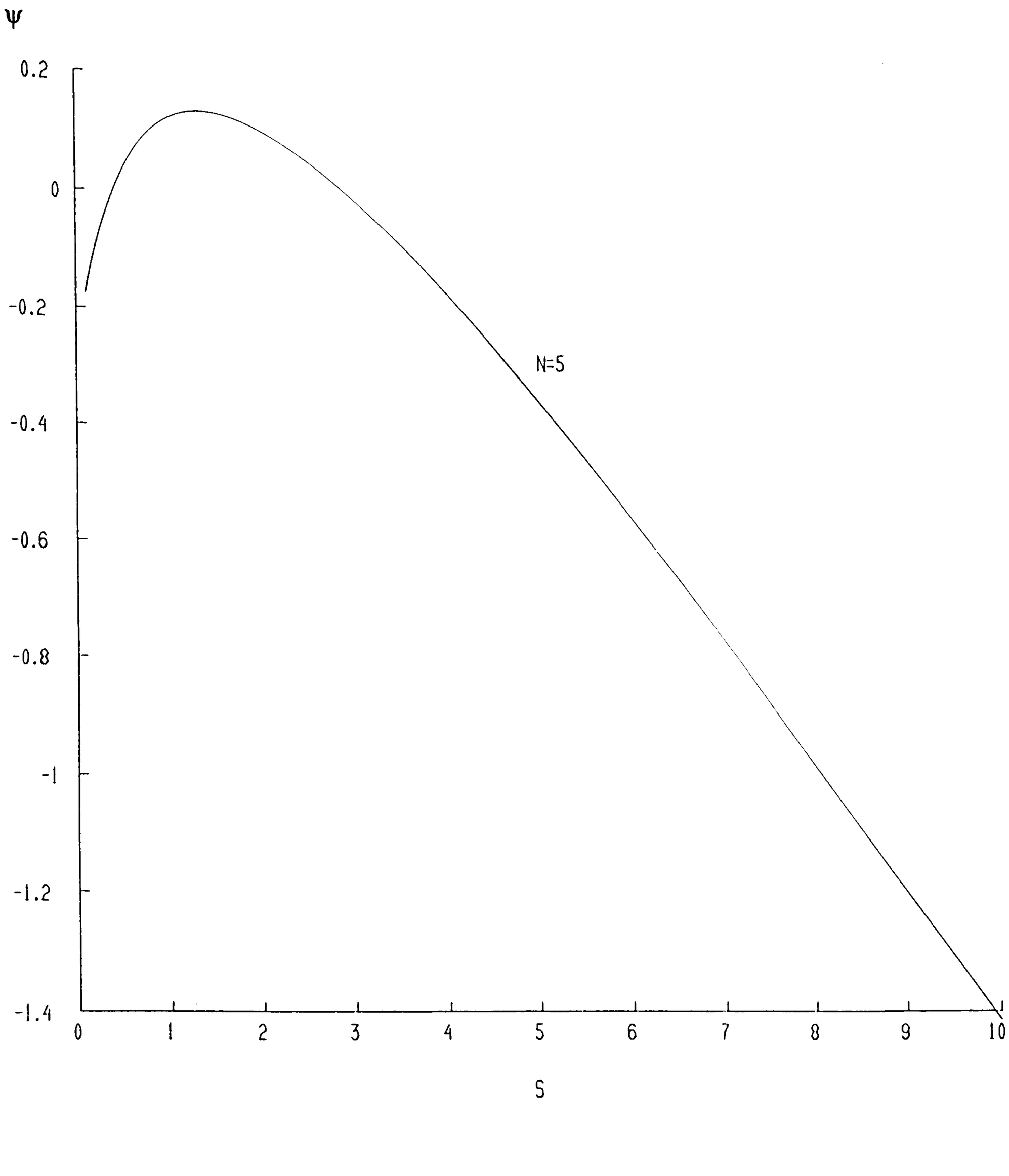


Figure 5.22. $\psi(s)$ for $\alpha_1 = 0.95$, $\alpha_2 = 0.05$.

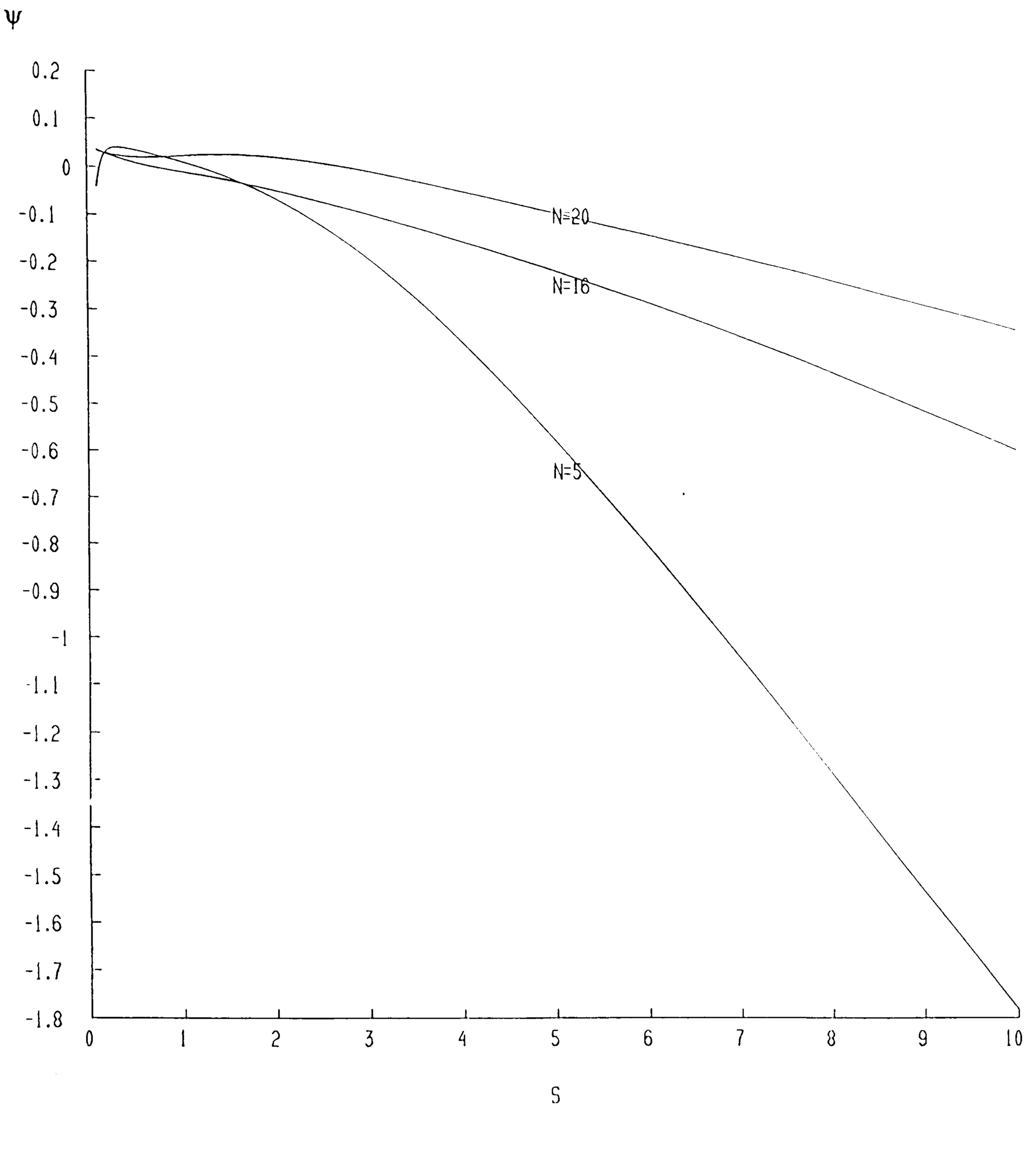
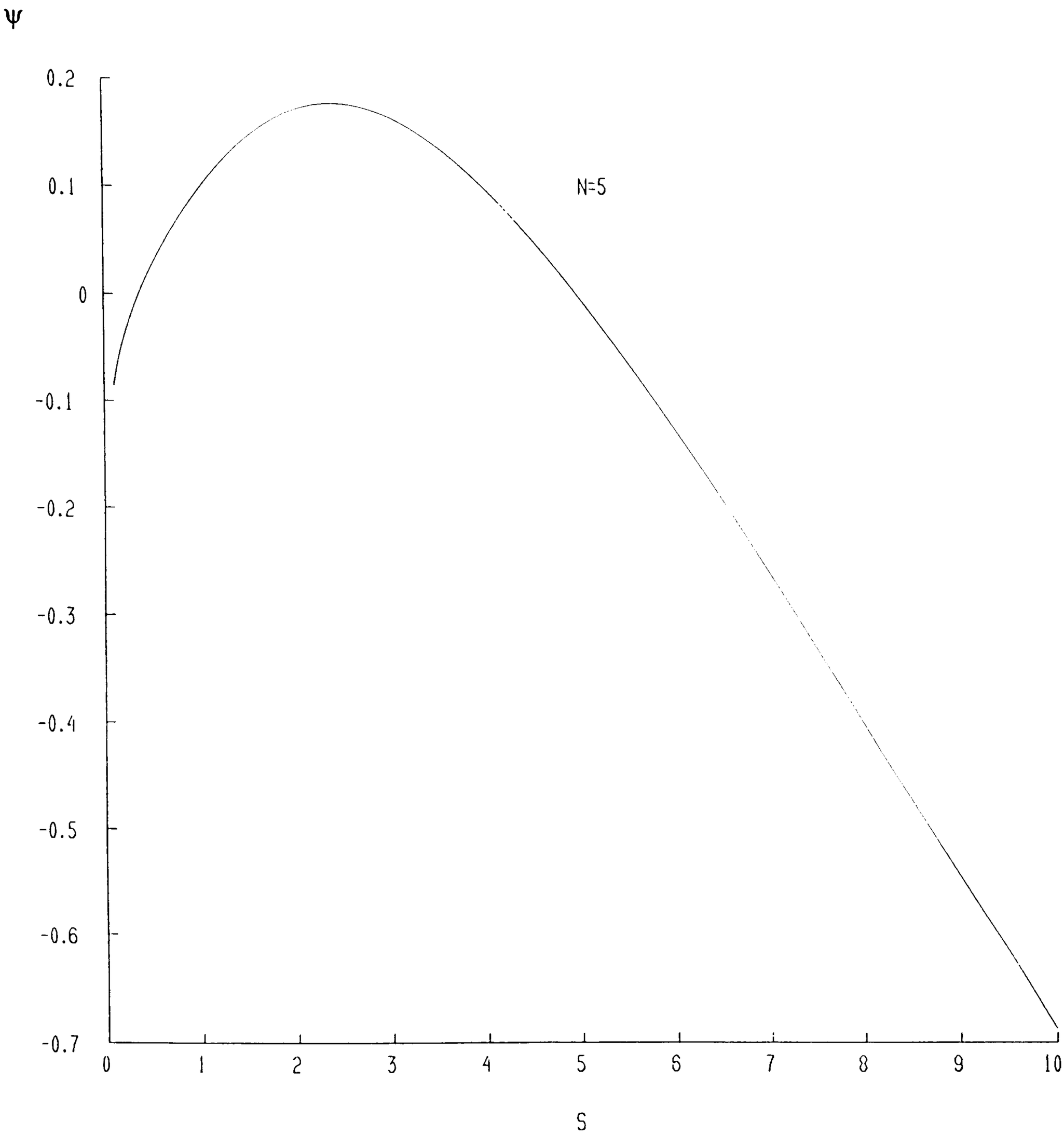


Figure 5.23. $\psi(s)$ for $\alpha_1 = 0.95$, $\alpha_2 = 0.45$.



5.5. Simulation Study of the Parameter Estimators.

For the purposes of this study we assume the initial dose to be known, choosing $\alpha_3 = 100$. Results are based on 1000 repetitions in each case. There are a number of distinct themes for us to investigate, beginning with a study of $\underline{\hat{\alpha}}_{ols}$ (see (5.71)). We collect data on the numerical algorithms employed, as well as the statistical performance indicators of M.S.E. and Relative Bias. Another feature to consider is the problem discussed earlier, when the system empties before sampling is complete. This we term a Range Failure, denoted by RFAIL. Recall that we employ the Gauss-Laguerre choice of sampling points throughout. A failure to converge of the NAG routines used for minimizing the least-squares function is referred to as an IFAIL. Note that the starting values used for OLS were the true parameter values.

5.5.1. Ordinary Least-Squares.

The results are summarized in Table 5.7 below, obtained using NAG routine E04JAF which is a simple non-derivative based algorithm. We reserve most discussion for later comparison with transform methods, but comment here that OLS appears to be very unstable. Before moving on we note that this results format is used throughout this simulation work, so the key to Table 5.7 may be universally applied. In the body of the table we use the E-notation to indicate powers of 10.

Key to Table 5.7:

MSE for $\hat{\alpha}_1$
MSE for $\hat{\alpha}_2$
Relative Bias for $\hat{\alpha}_1$
Relative Bias for $\hat{\alpha}_2$
RFAIL.IFAIL

(i) - Complete Failure of Sampling Scheme to Exist.

For each combination of parameter values and sample size, the entries, quoted correct to 5sf, are as follows: When an IFAIL or RFAIL is detected, the sample concerned is rejected and another is drawn. This process continues until 1000 samples have been collected, and it is for these samples that statistical performance has been assessed.

5.5.2. Two *s*-Values.

An additional feature now is the distribution of the optimal $\underline{s}$, chosen by the least-squares criterion (5.70). Our discussion of simulation results, given in Tables 5.8 and 5.8.1 below, is split into two distinct parts to reflect this. In the first we compare and contrast this study with that for OLS. Secondly, we discuss the observed distribution of optimal $\underline{s}$. In preliminary work the starting value $\underline{s} = (0.3, 0.05)'$ for NAG routine E04JAF seemed to work quite well. For this method then, we expect to observe an off-diagonal ($s_1 \neq s_2$) solution most of the time. However, we do have a particular interest in whether diagonal solutions are found on occasion, even for this starting-value. To this end, we give some representative plots of $|s_1 - s_2|$.

Turning first to a comparison with OLS, we note that the transform approach seems to perform much better numerically . This is manifest by the much reduced number of IFAILs. However, the case for $\alpha_1 = 0.95$ is incomplete due to a total failure of the NAG algorithm. Considering the relatively stable numerical performance for other parameter combinations, it it seems that this problem has less to do with the method of estimation than with the algorithm employed. As far as a statistical comparison is concerned, OLS performs better *when* it doesn't fail. However, considering its poor numerical performance, this seems to be little consolation.

With regards to the competing methods of transform construction, no obvious winner emerges. In terms of numerical stability there is little to choose, and the same conclusion applies to the statistical performance measures. There is one case of particular note though, where $\alpha_1 = 0.50$ and $\alpha_2 = 0.45$. We see that Gauss-Laguerre significantly out-performs its competitor, although the reason for this is not obvious.

Table 5.7. Simulation Results for OLS.

Intensities		Sample Size				
α_1	α_2	5	12	16	20	24
0.10	0.05	4.4922E-4	2.5993E-4	2.4364E-4	2.2762E-4	2.4680E-4
		3.7157E-4	4.8328E-5	3.8775E-5	3.6470E-5	3.3459E-5
		1.1385E-2	1.3908E-2	1.5716E-2	1.5956E-2	1.4784E-2
		-6.4181E-2	5.4470E-3	4.1410E-3	6.4735E-3	-5.1542E-4
		0.556	0.240	0.240	0.211	47.199
0.50	0.05	4.3682E-3	4.1483E-3	4.0153E-3	4.0120E-3	3.9522E-3
		7.0480E-5	4.0749E-5	3.5700E-5	3.4546E-5	3.3226E-5
		4.8103E-3	5.7021E-3	3.3633E-3	3.2737E-3	5.2926E-3
		1.0823E-2	1.1000E-2	1.2223E-2	1.0647E-2	3.8860E-3
		0.357	0.319	0.282	27.260	221.224
0.95	0.05	9.7921E-3	8.9262E-3	8.9095E-3	8.7004E-3	8.8810E-3
		6.8108E-5	3.9124E-5	3.5147E-5	3.4200E-5	3.1348E-5
		-6.8719E-2	-6.5005E-2	-6.4196E-2	-6.1844E-2	-6.4016E-2
		1.7665E-2	1.5296E-2	1.2970E-2	1.0728E-2	4.5388E-3
		0.1188	0.988	0.927	47.902	398.934
0.50	0.45	6.6354E-3	(i)	(i)	(i)	(i)
		2.0230E-3	(i)	(i)	(i)	(i)
		7.1582E-2	(i)	(i)	(i)	(i)
		-2.5958E-2	(i)	(i)	(i)	(i)
		488.834				
0.95	0.45	1.6089E-2	(i)	(i)	(i)	(i)
		2.8257E-3	(i)	(i)	(i)	(i)
		-9.2672E-2	(i)	(i)	(i)	(i)
		-7.1403E-3	(i)	(i)	(i)	(i)
		2396.792				
0.95	0.90	(i)	(i)	(i)	(i)	(i)

In this work we have a particular interest in the approach based on a single s-value, since this can lead to considerable simplifications, especially in numerical matters. To this end our discussion now focuses on the distribution of optimal s-values found by least-squares. In Figures 5.24-5.27 below we give representative plots of the quantity $|s_1 - s_2|$ for each method of transform construction in turn. Even though the pictorial results given are only for the case $\alpha_1 = 0.10$, $\alpha_2 = 0.05$ they do provide an accurate portrayal of the overall picture. We see that for the smaller samples sizes a diagonal solution is much more likely to be found. This does make some sense. Given that the starting values are in the vicinity of an off-diagonal solution, then we would expect this solution to be found, especially for the larger sample sizes. For the smaller sample sizes there must be a greater chance of finding a diagonal solution instead. Appendix II shows that there are up to four solutions to be detected by least-squares. There are two off-diagonal solutions, which are mirror images in the line $s_1 = s_2$. These are supplemented by as many as two diagonal solutions, corresponding to each of the two values of s. These solutions will only be detected if $\hat{L}'(s)$ matches $L'(s)$ in a region where $\hat{L}(s)$ matches $L(s)$ as well, as dictated by the results of Appendix II.

Table 5.8. Simulation Results for Gauss-Laguerre Transform Construction.

Intensities α_1 α_2	Sample Size 5	12	16	20	24
0.10 0.05	6.4988E-4	3.2300E-4	3.0513E-4	3.0215E-4	3.0553E-4
	1.1420E-3	6.1776E-5	3.9390E-5	3.2844E-5	2.9795E-5
	0.10326	1.6205E-2	1.4013E-2	1.1282E-2	1.0418E-2
	0.49595	4.2593E-2	2.4810E-2	1.5036E-2	5.9713E-3
	0.0	0.7	0.8	0.11	44.19
0.50 0.05	7.0570E-3	4.2796E-3	4.1521E-3	4.2262E-3	4.0872E-3
	1.8063E-4	4.1446E-5	3.5273E-5	3.1735E-5	3.0469E-5
	8.3901E-2	1.0297E-2	5.9929E-3	4.8699E-3	4.6343E-3
	0.20692	2.2961E-2	1.5021E-2	9.4255E-3	3.5653E-3
	0.1	0.9	0.10	24.6	187.10
0.95 0.05	*	*	*	*	*
0.50 0.45	7.5828E-3	(i)	(i)	(i)	(i)
	2.9707E-3	(i)	(i)	(i)	(i)
	2.6528E-2	(i)	(i)	(i)	(i)
	-3.2276E-4	(i)	(i)	(i)	(i)
	271.28				
0.95 0.45	*	(i)	(i)	(i)	(i)
0.95 0.90	(i)	(i)	(i)	(i)	(i)

* - Internal NAG error.

Table 5.8.1. Simulation Results for Riemann-Sum Transform Construction.

Intensities α_1 α_2	Sample Size 5	12	16	20	24
0.10 0.05	3.7227E-4	3.1049E-4	2.9890E-4	2.9700E-4	3.0067E-4
	3.6827E-4	6.9932E-5	4.5288E-5	3.7708E-5	3.2961E-5
	-7.9405E-2	-1.9569E-2	-6.4063E-3	-3.6626E-3	-1.2479E-3
	-0.25718	-6.3744E-2	-3.2500E-2	-1.8284E-2	-1.2770E-2
	0.5	0.2	0.6	0.7	44.14
0.50 0.05	8.2618E-3	4.0536E-3	3.9063E-3	3.9725E-3	3.9083E-3
	1.5343E-4	4.3212E-5	3.7694E-5	3.3332E-5	3.1323E-5
	-0.14856	-4.1293E-2	-2.8714E-2	-2.0712E-2	-1.6020E-2
	-0.19712	-7.0649E-3	3.2477E-3	6.2655E-3	3.2005E-5
	0.0	0.7	0.13	24.11	187.10
0.95 0.05	*	*	*	*	*
0.50 0.45	1.1301E-2	(i)	(i)	(i)	(i)
	3.8481E-3	(i)	(i)	(i)	(i)
	-0.14262	(i)	(i)	(i)	(i)
	-5.8774E-2	(i)	(i)	(i)	(i)
	271.5				
0.95 0.45	*	(i)	(i)	(i)	(i)
0.95 0.90	(i)	(i)	(i)	(i)	(i)

Figure 5.24. Distribution of $|s_1 - s_2|$ for Gauss-Laguerre Transform Construction, $n = 5$.

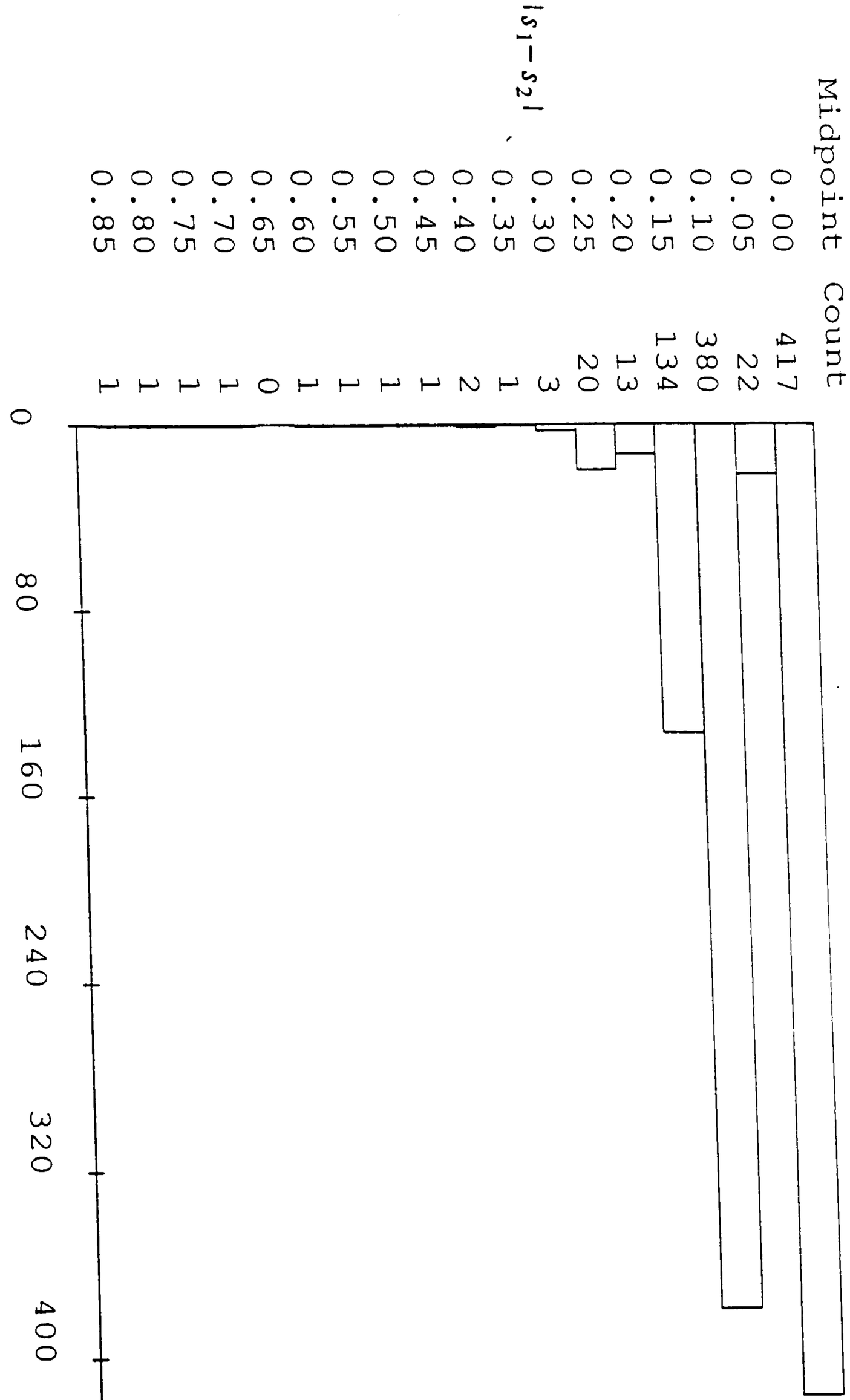


Figure 5.25. Distribution of $|s_1 - s_2|$ for Gauss-Laguerre Transform Construction, $n = 24$.

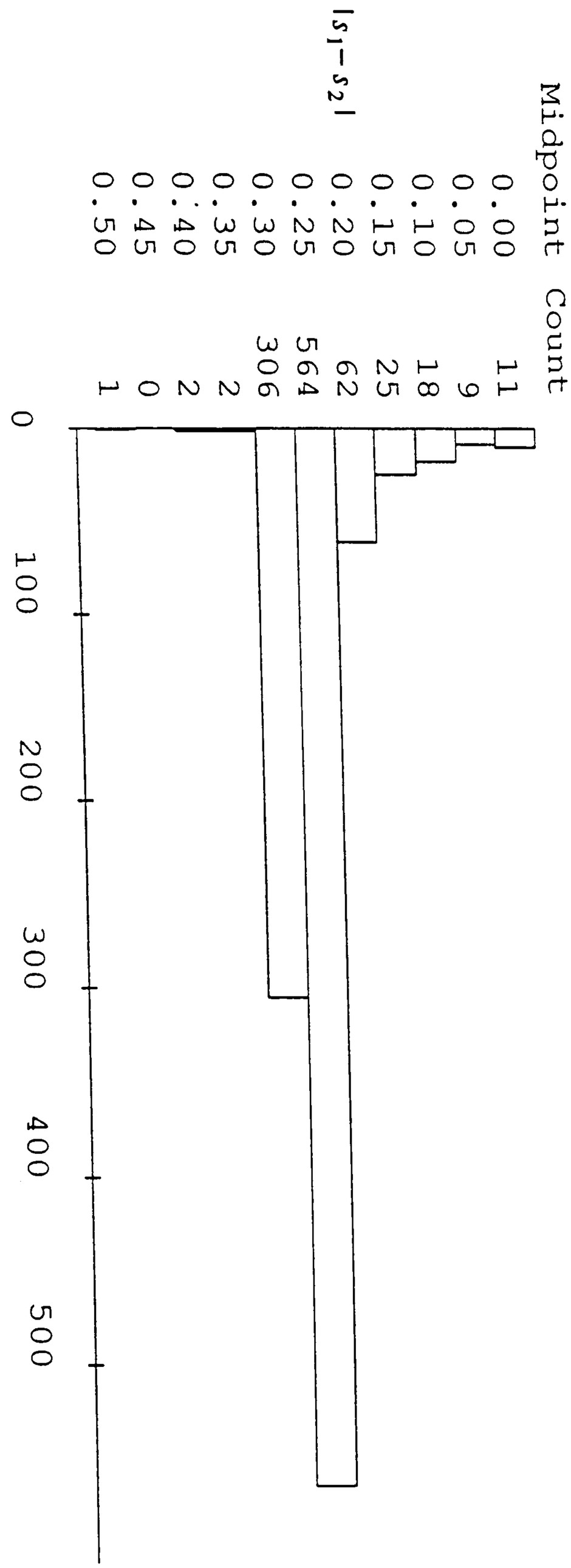


Figure 5.26. Distribution of $|s_1 - s_2|$ for Riemann-Sum Transform Construction, $n = 5$.

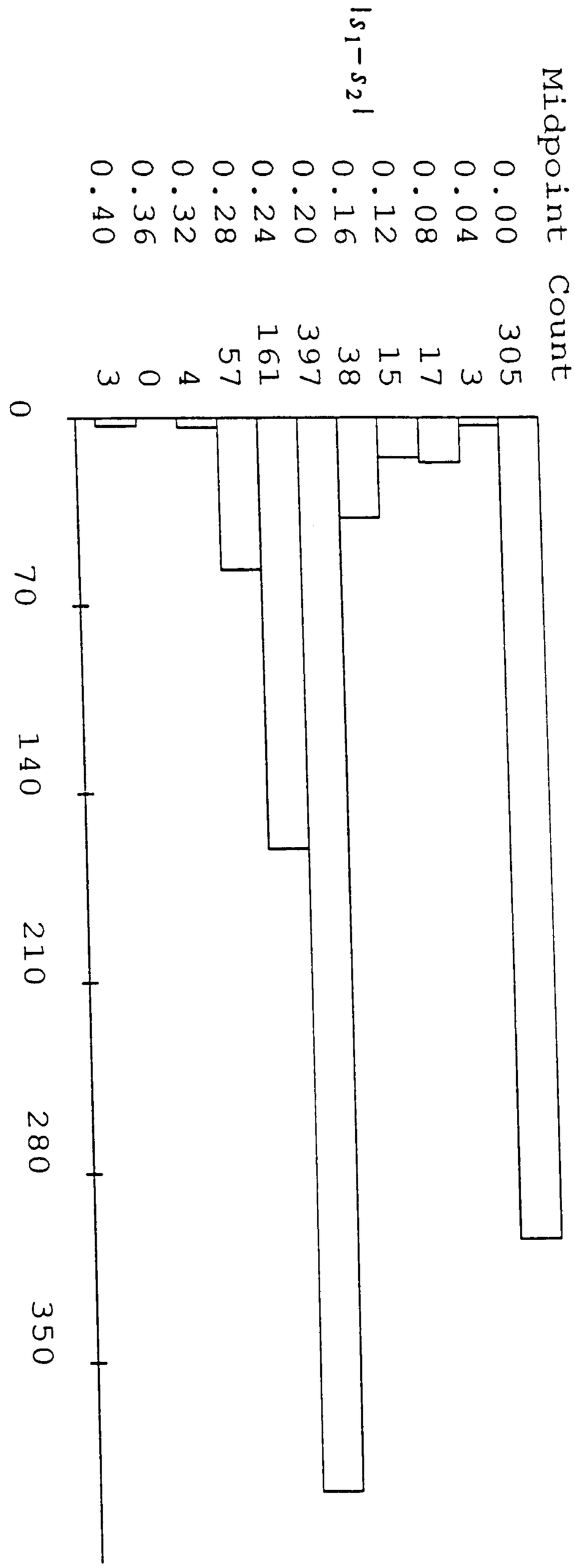
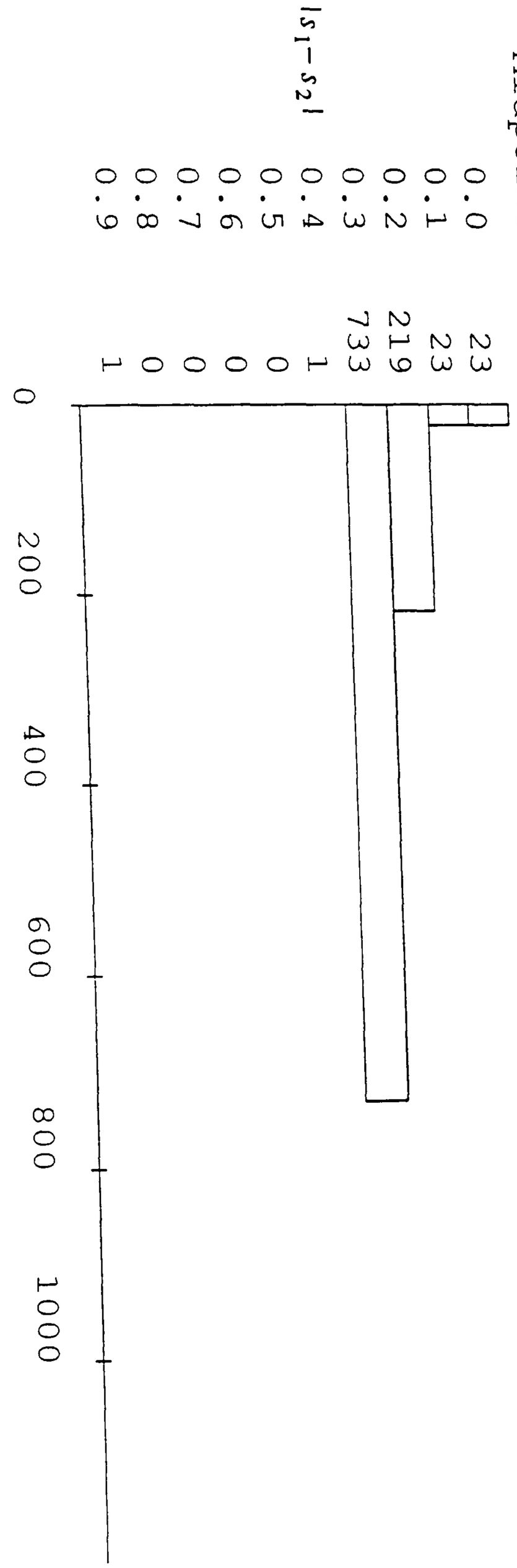


Figure 5.27. Distribution of $|s_1 - s_2|$ for Riemann-Sum Transform Construction, $n = 24$.



5.5.3. One *s*-Value.

The NAG routine E04JAF failed completely in this context, and was replaced by E04ABF. This algorithm is only suitable for one-dimensional minimization, and, like E04JAF, is not derivative-based. In the light of this it is difficult to make IFAIL comparisons with the earlier studies. However, for this study, summarized in Tables 5.9 and 5.9.1 below, there is little to choose between the methods of transform construction in terms of their IFAIL performance, the exception to this being for the larger-sample cases of $\alpha_1 = 0.10$, $\alpha_2 = 0.05$. For these combinations Gauss-Laguerre performs considerably worse. The picture is less clear in terms of MSE and Relative Bias where, once again, no winner is obvious.

In comparison with OLS, we see that the statistical performance is generally of the same order. Recall, however, that we only have to perform a one-dimensional search here. In addition, there are potential benefits in terms of robustness, a topic which is discussed in the next chapter. An interesting further point to note is the comparison of $MSE(\hat{\alpha}_1)$ and $MSE(\hat{\alpha}_2)$ for this study. Regardless of the method of construction, estimation of α_2 is uniformly better than for α_1. This seems to be a general rule for the other studies as well.

We are now in a position to compare the one and two *s*-value approaches. The most obvious difference in performance is for $\underline{\alpha} = (0.10, 0.05)'$, where the two s-value method is much better, and particularly so as the sample size increases. The explanation for this might be that the sample size is made better use of in the two s-value approach. Apart from this there is little to choose, and when a method does perform better, the difference is largely marginal in nature.

Overall, it seems that we do not unduly detract from performance by employing just one *s*-value rather than two. In practice, it seems that graphical procedures would be sufficient to locate an optimal value of *s*, reducing numerical problems correspondingly. We move on now to consider the distribution of the optimal *s*, s_{opt} say, found in this study. There are two issues of central importance here. First, where s_{opt} tends to occur and, secondly, whether its location depends on the method of construction. We consider these questions case-by-case below.

Table 5.9. Simulation Results for Gauss-Laguerre Transform Construction.

Intensities α_1 α_2	Sample size 5	12	16	20	24
0.10 0.05	5.7109E-4	2.9543E-4	2.5942E-4	2.6996E-4	2.6498E-4
	7.2439E-4	5.3389E-5	4.3622E-5	3.8218E-5	3.5310E-5
	4.0943E-2	2.2815E-2	1.8407E-2	2.1386E-2	1.5605E-2
	0.14927	2.5124E-2	8.6780E-3	8.5425E-3	6.4989E-3
	0.3	0.13	0.58	0.90	42.121
0.50 0.05	4.6250E-3	4.3114E-3	4.3634E-3	4.5175E-3	4.3828E-3
	9.1784E-5	4.1574E-5	3.8805E-5	3.7955E-5	3.4059E-5
	1.7344E-2	9.6310E-3	7.2666E-3	9.4895E-3	7.6029E-3
	5.1849E-2	1.1135E-2	1.1214E-2	7.5231E-3	4.3619E-3
	0.4	0.23	0.26	20.27	175.19
0.95 0.05	1.6457E-2	1.8481E-2	1.8634E-2	1.8953E-2	1.8205E-2
	7.9152E-5	4.2844E-5	4.0242E-5	3.8722E-5	3.6525E-5
	2.0151E-2	1.2733E-2	1.1587E-2	1.1851E-2	9.6374E-3
	3.8228E-2	9.7428E-2	9.1980E-2	7.0700E-3	3.3504E-3
	0.6	0.14	0.10	25.8	196.12
0.50 0.45	8.3001E-3	(i)	(i)	(i)	(i)
	3.2671E-3	(i)	(i)	(i)	(i)
	8.3162E-3	(i)	(i)	(i)	(i)
	5.4613E-3	(i)	(i)	(i)	(i)
	210.9				
0.95 0.45	2.7260E-2	(i)	(i)	(i)	(i)
	2.9996E-3	(i)	(i)	(i)	(i)
	1.8700E-2	(i)	(i)	(i)	(i)
	-3.7780E-3	(i)	(i)	(i)	(i)
	1131.9				
0.95 0.90	(i)	(i)	(i)	(i)	(i)

Table 5.9.1. Simulation Results for Riemann-Sum Transform Construction.

Intensities α_1 α_2	Sample Size 5	12	16	20	24
0.10 0.05	4.1432E-4	2.7704E-4	2.5735E-4	2.6075E-4	2.6071E-4
	3.9043E-4	5.5494E-5	4.4790E-5	3.6763E-5	2.6316E-5
	-5.4110E-2	-1.5445E-2	-3.3087E-3	8.0265E-3	7.5069E-3
	-0.14456	-2.9799E-2	-1.1447E-2	1.7293E-3	2.9445E-3
	0.12	0.6	0.10	0.23	42.45
0.50 0.05	5.5765E-3	4.0519E-3	4.0969E-3	4.2810E-2	4.1307E-3
	9.0018E-5	4.3069E-5	3.9274E-5	3.8505E-5	3.4070E-5
	-8.8160E-2	-3.1631E-2	-1.7794E-2	-9.8720E-3	-8.9720E-3
	-3.3550E-2	9.0484E-3	1.4068E-2	1.0248E-2	6.2396E-3
	0.5	0.11	0.19	20.20	175.21
0.95 0.05	2.1478E-2	1.7347E-2	1.7348E-2	1.7469E-2	1.6876E-2
	7.7172E-5	4.4521E-5	4.1227E-5	3.9433E-5	3.6884E-5
	0.10256	-4.4137E-2	-2.8370E-2	-2.0794E-2	-1.7827E-2
	1.2356E-2	1.5580E-2	1.4663E-2	1.1517E-2	6.0473E-3
	0.4	0.8	0.4	25.12	196.9
0.50 0.45	8.4157E-3	(i)	(i)	(i)	(i)
	3.5281E-3	(i)	(i)	(i)	(i)
	-6.8596E-2	(i)	(i)	(i)	(i)
	4.7586E-3	(i)	(i)	(i)	(i)
	210.4				
0.95 0.45	2.7421E-2	(i)	(i)	(i)	(i)
	3.2663E-3	(i)	(i)	(i)	(i)
	-5.8924E-2	(i)	(i)	(i)	(i)
	1.0044E-2	(i)	(i)	(i)	(i)
	1131.11				
0.95 0.90	(i)	(i)	(i)	(i)	(i)

(1) $\alpha_1 = 0.10, \alpha_2 = 0.05.$

We denote the optimal s found using Gauss-Laguerre and Riemann-Sum transform construction by sG and sR respectively. The results for this case are given below, where we quote the mean and median as summary statistics.

Table 5.10. Distribution of s_{opt} for $\underline{\alpha} = (0.10, 0.05)'$.

	sG		sR	
n	mean	median	mean	median
5	0.37	0.35	0.27	0.24
24	0.23	0.13	0.23	0.12

There is an essentially smooth transition between these two cases. It is clear from this summary, as well as a plot of $|sG - sR|$, that the optima are found in different regions of s-space. As sample size increases the optima move closer together, as we might expect. Note that a few simulations resulted in s_{opt} escaping to the boundary of the region of search.

(2) $\alpha_1 = 0.50, \alpha_2 = 0.05.$

For $n = 5$ only one observation escaped the region of search. In contrast, escapees were 11 (sG) and 14 (sR) for $n = 24$. There is a notable shift upwards in the distribution of s_{opt} for Riemann-Sum transform construction, when compared to Table 5.10.

Table 5.11. Distribution of s_{opt} for $\underline{\alpha} = (0.50, 0.05)'$.

	sG		sR	
n	mean	median	mean	median
5	0.38	0.35	0.39	0.36
24	0.28	0.16	0.28	0.15

(3) $\alpha_1 = 0.95, \alpha_2 = 0.05$.

Escapees increase in the $n = 24$ case, to 38 (sG) and 34 (sR).

Table 5.12. Distribution of s_{opt} for $\underline{\alpha} = (0.95,0.05)'$.

	sG		sR	
n	mean	median	mean	median
5	0.41	0.37	0.45	0.40
24	0.33	0.20	0.32	0.19

For the remaining two cases we only have information for $n = 5$, so we consider them together :

Table 5.13. Distribution of s_{opt} for Remaining Cases.

	sG		sR		
Intensities	mean	median	mean	median	escapees (sG, sR)
(0.50,0.45)′	0.34	0.34	0.48	0.48	4, 40
(0.95,0.45)′	0.45	0.46	0.46	0.67	5, 109

We see a large increase in the number of escapees, but only for the Riemann-Sum method.

In general, s_{opt} tends to increase as the intensities increase. The distribution of s_{opt} is almost uniformly positively-skewed, with this skewness dying-out as the αs increase for the quadrature-based method. For the Riemann-Sum technique we see negative-skewness emerge, although there are signs of this for quadrature as well. One final point is that the simulations were completed much more quickly than for the two s-value approach. The saving in time was in the region of 50%.

We conclude the simulation work of this chapter by returning to a problem touched on in a number of places earlier. Many of the simulations could not be carried-out because the sampling schemes did not exist, due to material leaving the system before sampling is complete. In order to

study the larger intensities we must, therefore, increase the initial dose α_3. In section 5.6 we show that $N_2(t) \sim Bin\,(\alpha_3,\ n_2(t)/\alpha_3)$, which we make use of now to suggest a reasonable value of α_3. The most difficult case is $\underline{\alpha} = (0.95, 0.90)'$, so an α_3 which improves the RFAILs here will naturally lead to improvements for less awkward cases. The criterion we apply is

$$Pr\,(\,N_2(t_{\max}\,) \geq 1\,) \geq 0.90\,,$$

where $t_{\max}$ denotes the largest sampling point. Thus,

$$1 - [1 - 19(e^{-0.9t_{\max}} - e^{-0.95t_{\max}}\,)]^{\alpha_3} \geq 0.90\,.$$

Recall that we choose the sampling points $\underline{t}$ by quadrature, so $t_{\max} \approx 82$. We find

$$\alpha_3 \geq \frac{\log(0.10)}{\log[1 - 19(e^{-73.8} - e^{-77.9})]}\,,$$

and making use of the approximation $log\,(1-x) \approx -x$, for small $x > 0$, yields

$$\alpha_3 \geq \frac{-log\,(0.10)}{19\,(e^{-73.8} - e^{-77.9})} \approx 1.39 * 10^{31}\,,$$

a rather large number!

It is not really practical for us to simulate the OCO model for such a large initial dose, so we look at the largest case attempted: $\alpha_3 = 1000$. The results are presented in Tables 5.14 and 5.14.1 below for the estimators based on a single value of s. Performance is generally improved, and we note that estimation of α_2 is still superior to that of α_1. Of particular note is Relative Bias, where quadrature performs much better. RFAILs are much reduced, but we are still unable to extend the simulation study any further than was possible for $\alpha_3 = 100$. In terms of numerical stability and the distribution of s_{opt}, our previous conclusions hold unaltered. Clearly the larger the initial dose, the better the information we are able to obtain on transition intensities, at least over the range considered here. We are now in a position to conduct a more theoretical investigation, which follows after a summary of the key conclusions of the simulation work presented here.

5.6. Summary of Simulation Results.

In this chapter we have conducted quite a detailed simulation study of the estimators proposed. We now present a distillation of the most important results, which fall into the broad

categories of numerical and statistical performance.

Our benchmark has been ordinary least-squares, which demonstrated good statistical properties in terms of relative bias and mean-squared-error. However, its numerical performance was poor when compared to transforms, which have the further important advantage that the estimators so obtained take explicit forms. At the very least, then, transforms could be used to provide starting values for more complex procedures.

When employing transforms in estimation, we have the potential advantage of dimension reduction. The basic estimating equations require us to select p s-values when there are p unknown parameters to estimate. We investigated these estimators in sub-section 5.2 of this work. However, Appendix II shows us that we obtain well-defined estimators by allowing the s-values to approach some common value, s say. The results of sub-section 5.2 suggested that the least-squares criterion for selecting $\underline{s}$ was detecting such a "diagonal" solution. In sub-section 5.3 we concentrated on investigating this solution, and showed that it performed well when compared to the approach based on two values of s. This effectively reduces a two-dimensional search to a line-search. Dimension reduction is an important feature of transform estimators, which is returned to in the efficiency work to follow.

Since there are competing methods of transform construction, a comparison between them has been a key theme of the work so far, and will continue to be so. With only very few exceptions, the quadrature and Riemann-sum techniques have performed broadly the same. It was noted, however, that the s-values selected by each method were not identical. Indeed, they were sometimes very far apart.

Table 5.14. Simulation Results for $\alpha_3 = 1000$; Transform Construction by Quadrature.

Intensities α_1	Intensities α_2	Sample Size 5	12	16	20	24
0.10	0.05	6.5274E-5	2.9489E-5	2.6630E-5	2.4671E-5	2.4399E-5
		1.2458E-4	6.8095E-6	4.2322E-6	3.5608E-6	3.2165E-6
		1.3417E-2	6.7729E-3	4.7247E-3	2.7338E-3	2.0613E-3
		8.5913E-2	1.6014E-2	7.9022E-3	4.2735E-3	2.4763E-3
		0.4	0.7	0.19	0.40	0.65
0.50	0.05	4.6059E-4	4.2445E-4	4.4490E-4	4.3594E-4	4.4045E-4
		1.1278E-5	3.9936E-6	3.5910E-6	3.4210E-6	3.4019E-6
		5.2829E-3	1.4602E-3	5.3413E-4	-5.7356E-5	-1.2013E-4
		2.4121E-2	4.1292E-3	1.9598E-3	2.1616E-3	1.7796E-3
		0.4	0.14	0.18	0.18	0.14
0.95	0.05	1.6546E-3	1.6741E-3	1.8059E-3	1.8180E-3	1.8540E-3
		8.4598E-6	4.1584E-6	3.8589E-6	3.7945E-6	3.7812E-6
		4.7687E-3	1.5487E-3	8.7000E-4	2.1471E-4	1.2674E-4
		1.4414E-2	2.5139E-3	2.1980E-3	1.8699E-3	2.0526E-3
		0.4	0.14	0.10	0.7	0.5
0.50	0.45	7.6228E-4	(i)	(i)	(i)	(i)
		2.9937E-4	(i)	(i)	(i)	(i)
		2.7239E-3	(i)	(i)	(i)	(i)
		2.4451E-3	(i)	(i)	(i)	(i)
		0.4				
0.95	0.45	2.5860E-3	(i)	(i)	(i)	(i)
		2.9134E-4	(i)	(i)	(i)	(i)
		5.3951E-4	(i)	(i)	(i)	(i)
		1.4689E-3	(i)	(i)	(i)	(i)
		2.2				
0.95	0.90	(i)	(i)	(i)	(i)	(i)

Table 5.14.1. Simulation Results for $\alpha_3 = 1000$; Riemann-Sum Transform Construction.

Intensities α_1 α_2	Sample Size 5	12	16	20	24
0.10 0.05	7.7748E-5	3.3350E-5	2.7358E-5	2.5748E-5	2.4534E-5
	6.9954E-5	8.0286E-6	4.5983E-6	3.8243E-6	3.3585E-6
	-5.8949E-2	-2.1157E-2	-1.3443E-2	-8.9757E-3	-5.7559E-3
	-0.10015	-1.9121E-2	-7.8668E-3	-2.7353E-3	-1.3728E-4
	0.6	0.4	0.8	0.11	0.10
0.50 0.05	2.1885E-3	7.8118E-4	6.0282E-4	5.3179E-4	5.0943E-4
	1.1896E-5	4.3020E-6	3.7790E-6	3.4735E-6	3.4399E-6
	-8.5048E-2	-4.1022E-2	-2.8195E-2	-2.0650E-2	-1.6377E-2
	-6.3229E-3	-4.0180E-4	2.2879E-3	3.9181E-3	4.0065E-3
	0.5	0.2	0.2	0.6	0.18
0.95 0.05	1.0817E-2	4.6329E-3	3.3466E-3	2.7778E-3	2.4888E-3
	8.9347E-6	4.5808E-6	3.0526E-6	3.8932E-6	3.7160E-6
	-0.10301	-6.0445E-2	-4.2689E-2	-3.2485E-2	-2.6393E-2
	2.5620E-2	3.8677E-3	5.5032E-3	5.0900E-3	5.0771E-3
	0.4	0.3	0.5	0.7	0.11
0.50 0.45	2.0391E-3	(i)	(i)	(i)	(i)
	4.0987E-4	(i)	(i)	(i)	(i)
	-7.1618E-2	(i)	(i)	(i)	(i)
	1.1752E-2	(i)	(i)	(i)	(i)
	0.9				
0.95 0.45	6.9486E-3	(i)	(i)	(i)	(i)
	4.6140E-4	(i)	(i)	(i)	(i)
	-7.1095E-2	(i)	(i)	(i)	(i)
	2.6136E-2	(i)	(i)	(i)	(i)
	2.12				
0.95 0.90	(i)	(i)	(i)	(i)	(i)

5.7. Second-Order Properties.

We move away now from the simulation approach, and towards a more theoretical discussion. We first consider measures of error for the parameter estimators, which subsequently requires us to find the dispersion matrix of the observations. We discuss two approaches to the latter problem, and demonstrate an appealing intuitive solution.

5.7.1. Parameter Estimators.

We begin by finding the variance matrix of the transforms themselves. Recall that in the estimation setting there will be as many transforms as unknown parameters. Now, in the general form (5.73) we have

$$\begin{aligned} \hat{L}(s) &= \sum_{k=1}^{n} h_k(s) N_2(t_k) , \\ &= \underline{h}'(s)\underline{N}_2 \ , \ say, \end{aligned} \tag{5.79}$$

where

$$\underline{h}(s) = (h_1(s), \ldots, h_k(s))'$$

and

$$\underline{N}_2 = (N_2(t_1), \ldots, N_2(t_n))' .$$

Then

$$\begin{aligned} cov(\hat{L}_i, \hat{L}_j) &= cov[\sum_k h_k(s_i) N_2(t_k), \sum_l h_l(s_j) N_2(t_l)] \quad i,j=1,\ldots,p \\ &= \sum_{k,l} h_k(s_i) h_l(s_j) cov(N_2(t_k), N_2(t_l)) \\ \Rightarrow \ cov(\hat{L}_i, \hat{L}_j) &= \underline{h}'(s_i) var(\underline{N}_2) \underline{h}(s_j) . \end{aligned} \tag{5.80}$$

Bringing all of the transforms together, we obtain

$$var(\underline{\hat{L}}) = H'(\underline{s}) var(\underline{N}_2) H(\underline{s}) , \tag{5.81}$$

where

$$[H(\underline{s})]_{ij} = h_i(s_j) \quad i=1,\ldots,n, \ \ j=1,\ldots,p.$$

Note that the single s-value approach is easily included as a special case.

We can obtain an alternative expression for $var(\underline{\hat{L}})$, involving the dispersion matrix of the estimators, as follows. The estimating equations are

$$\hat{L}(s_i) = L(s_i; \underline{\hat{\alpha}}) \quad i=1,...,p,$$

and applying the delta-method to the above equations we find

$$var(\hat{L}_i) \approx \sum_{k,l} cov(\hat{\alpha}_k, \hat{\alpha}_l) \frac{\partial L_i}{\partial \alpha_k} . \frac{\partial L_i}{\partial \alpha_l} \tag{5.82}$$

$$\Rightarrow \quad var(\underline{\hat{L}}) \approx \nabla' var(\underline{\hat{\alpha}}) \nabla \, , \tag{5.83}$$

where

$$(\nabla)_{ij} = \frac{\partial L_j}{\partial \alpha_i} \quad i,j=1,...,p.$$

We may therefore write-down the approximation

$$H'(\underline{s}) var(\underline{N}_2) H(\underline{s}) \approx \nabla' var(\underline{\hat{\alpha}}) \nabla \, , \tag{5.84}$$

and upon writing

$$\kappa(\underline{s}) = H(\underline{s}) \nabla^{-1}$$

we obtain

$$var(\underline{\hat{\alpha}}) \approx \kappa'(\underline{s}) var(\underline{N}_2) \kappa(\underline{s}) \, . \tag{5.85}$$

So to calculate measures of error for our transform estimators we require the form of $var(\underline{N}_2)$. This is the topic of the next section.

5.7.2. The Observations.

We adopt a similar approach to that of Matis & Hartley (1971), although they worked in a rather more general setting, defining

$$P_{n_1,n_2}(t) = pr[N_1(t) = n_1, N_2(t) = n_2] \tag{5.86}$$
$$= P_{\underline{n}}(t), \; say.$$

Given that the transitions are assumed to follow Poisson processes, we have that the probability of a particle leaving the ith compartment in a time δt is

$$Pr[N_i(t+\delta t) = N_i(t) - 1] = \alpha_i n_i \, \delta t + o(\delta t) \quad i=1,2. \tag{5.87}$$

We find the forward equations in the usual way as

$$P_{\underline{n}}(t) = P_{\underline{n}}(t)[1-(\alpha_1 n_1+\alpha_2 n_2)\delta t]+P_{n_1,n_2+1}(t)\alpha_2(n_2+1)\delta t+P_{n_1+1,n_2-1}(t)\alpha_1(n_1+1)\delta t+o(\delta t)\,. \tag{5.88}$$

Simplifying and taking the limit $\delta t\rightarrow 0$, we obtain

$$\frac{dP_{n_1,n_2}(t)}{dt} = -(\alpha_1 n_1+\alpha_2 n_2)P_{\underline{n}}(t)+\alpha_2(n_2+1)P_{n_1,n_2+1}(t)+\alpha_1(n_1+1)P_{n_1+1,n_2-1}(t)\,. \tag{5.89}$$

This expression is valid for $\{n_1 = 0,1,...,\alpha_3-1;\, n_2 = 1,2,...,\alpha_3-1 : n_1+n_2\leq\alpha_3\}$, so we must take account of the special cases (a) $(n_1,0)$, $n_1 = 0,1,...,\alpha_3$ and (b) $(0,\alpha_3)$.

Case(a).

$$P_{n_1,0}(t+\delta t) = P_{n_1,0}(t)[1-\alpha_1 n_1\delta t]+P_{n_1,1}(t)\alpha_2\delta t+o(\delta t)$$

$$\Rightarrow \frac{dP_{n_1,0}(t)}{dt} = -\alpha_1 n_1 P_{n_1,0}(t)+\alpha_2 P_{n_1,1}(t)\,. \tag{5.90}$$

Case(b).

$$P_{0,\alpha_3}(t) = P_{0,\alpha_3}(t)[1-\alpha_2\alpha_3\delta t]+P_{1,\alpha_3-1}(t)\alpha_1\delta t+o(\delta t)$$

$$\Rightarrow \frac{dP_{0,\alpha_3}(t)}{dt} = -\alpha_2\alpha_3 P_{0,\alpha_3}(t)+\alpha_1 P_{1,\alpha_3-1}(t)\,. \tag{5.91}$$

We have now considered every combination of n_1 and n_2. To solve the differential equations (5.89-91) we introduce the bivariate probability generating function

$$G(u,v) = \sum_{\underline{n}} u^{n_1}v^{n_2}P_{\underline{n}}(t)\,. \tag{5.92}$$

In the first differential equation (5.89), we find

$$\frac{\partial}{\partial t}[u^{n_1}v^{n_2}P_{\underline{n}}(t)] = -(\alpha_1 n_1+\alpha_2 n_2)u^{n_1}v^{n_2}P_{\underline{n}}(t)+\alpha_2 u^{n_1}\frac{\partial}{\partial v}[v^{n_2+1}P_{n_1,n_2+1}(t)]+$$
$$\alpha_1 v\frac{\partial}{\partial u}[u^{n_1+1}v^{n_2-1}P_{n_1+1,n_2-1}(t)]\,.$$

Continuing in this fashion, we obtain

$$\frac{\partial}{\partial t}[u^{n_1}v^{n_2}P_{\underline{n}}(t)] = -\alpha_1 u\frac{\partial}{\partial u}[u^{n_1}v^{n_2}P_{\underline{n}}(t)]-\alpha_2 v\frac{\partial}{\partial v}[u^{n_1}v^{n_2}P_{\underline{n}}(t)]+$$

$$\alpha_2\frac{\partial}{\partial v}[u^{n_1}v^{n_2+1}P_{n_1,n_2+1}(t)]+\alpha_1 v\frac{\partial}{\partial u}[u^{n_1+1}v^{n_2-1}P_{n_1+1,n_2-1}(t)]\,. \tag{5.93}$$

Filling-in the special cases and taking $\sum_{\underline{n}}$ yields

$$\frac{\partial G(u,v)}{\partial t} = \alpha_1(v-u)\frac{\partial G(u,v)}{\partial u} + \alpha_2(1-v)\frac{\partial G(u,v)}{\partial v} . \tag{5.94}$$

So to find an expression for the p.g.f. (5.92) we must solve for G in the partial differential equation

$$\frac{\partial G}{\partial t} + \alpha_1(u-v)\frac{\partial G}{\partial u} + \alpha_2(v-1)\frac{\partial G}{\partial v} = 0 . \tag{5.95}$$

To do this we must first obtain two first integrals, $c_1(u,v,t)$ and $c_2(u,v,t)$ say, from the auxiliary equations

$$\frac{dt}{1} = \frac{du}{\alpha_1(u-v)} = \frac{dv}{\alpha_2(v-1)} \quad (= \frac{dG}{0}) . \tag{5.96}$$

c_1 follows readily from

$$\frac{dt}{1} = \frac{dv}{\alpha_2(v-1)}$$

$$\Rightarrow\ e^{-\alpha_2 t} = c_1(v-1)$$

$$\Rightarrow\ c_1(u,v,t) = (v-1)^{-1}e^{-\alpha_2 t} .$$

It is easily verified, by substitution in (5.95), that c_1 is indeed a first integral. We obtain a second first integral from

$$\frac{dt}{1} = \frac{du}{\alpha_1(u-v)}$$

$$\Rightarrow\ \frac{du}{dt} = \alpha_1(u-v) .$$

We can eliminate v by using $c_1 = (v-1)^{-1}e^{-\alpha_2 t}$ to give

$$\frac{du}{dt} = \alpha_1 u - \alpha_1 - \alpha_1 c_1^{-1} e^{-\alpha_2 t} ,$$

which may be expressed in the form

$$\frac{du}{dt} + B(t)u = Q(t)$$

where $B(t) = -\alpha_1$ and $Q(t) = -\alpha_1[1 + c_1^{-1}e^{-\alpha_2 t}]$. Employing the method of integrating factors, we find

$$u = 1 + c_2 e^{\alpha_1 t} + \frac{c_1^{-1}\alpha_1 e^{-\alpha_2 t}}{\alpha_1 + \alpha_2} .$$

Substituting for c_1, we obtain the second first integral

$$c_2(u,v,t) = [(u-1)-\alpha_1(\alpha_1+\alpha_2)^{-1}(v-1)]e^{-\alpha_1 t} .$$

Again, it is straightforward to verify that c_2 satisfies (5.95).

However, these are not the only first integrals. By inspection of these solutions, we also have

$$c_1 = (v-1)e^{-\alpha_2 t}$$

and

$$c_2 = e^{-\alpha_1 t}[u-1-\alpha_1(\alpha_1-\alpha_2)^{-1}(v-1)] .$$

These solutions are a little more convenient for finding the particular solution to our problem, so they are employed as follows.

The general solution to (5.95) is

$$G(u,v,t) = g\{(v-1)e^{-\alpha_2 t}, e^{-\alpha_1 t}[u-1-\alpha_1(\alpha_1-\alpha_2)^{-1}(v-1)]\} , \tag{5.97}$$

for some function $g \in C^1$ to be found from the boundary condition

$$G(u,v,0) = u^{\alpha_3} .$$

That is,

$$g\{v-1, u-1-\alpha_1(\alpha_1-\alpha_2)^{-1}(v-1)\} = u^{\alpha_3} .$$

Let $l_1 = (v-1)e^{-\alpha_1 t}$ and $l_2 = [u-1-\alpha_1(\alpha_1-\alpha_2)^{-1}(v-1)]e^{-\alpha_1 t}$, so that

$$g(l_1,l_2) = [1+l_2+\alpha_1(\alpha_1-\alpha_2)^{-1}l_1]^{\alpha_3}$$

$$\Rightarrow\ G(u,v,t) = \{1+[u-1-\alpha_1(\alpha_1-\alpha_2)^{-1}(v-1)]e^{-\alpha_1 t}+\alpha_1(\alpha_1-\alpha_2)^{-1}(v-1)e^{-\alpha_2 t}\} .$$

On writing

$$p_1(t) = e^{-\alpha_1 t} \text{ and } p_2(t) = \frac{\alpha_1}{\alpha_1-\alpha_2}(e^{-\alpha_2 t}-e^{-\alpha_1 t}) \tag{5.98}$$

we find the solution

$$G(u,v,t) = [1+(u-1)p_1(t)+(v-1)p_2(t)]^{\alpha_3} . \tag{5.99}$$

It is easily verified that this does satisfy (5.95), as required.

Upon setting $u = 1$, we obtain

$$G(v,t) = [1+(v-1)p_2(t)]^{\alpha_3}$$
$$\Rightarrow\ N_2(t) \sim Bin(\alpha_3, p_2(t)),$$

so that $var(N_2(t)) = \alpha_3 p_2(t)(1-p_2(t))$. The results so far do not give information on the observations collectively. As such, there doesn't seem to be any obvious way to obtain $cov(N_2(t_1), N_2(t_2)$ for $t_2 > t_1$. We go now to look at an appealing technique which gives intuitive access to the above distributional result, as well as the dispersion matrix of the observations, $\underline{N}_2$.

We define a vector $\underline{p}(t) = (p_1(t), p_2(t))'$ of compartmental occupancy probabilities. Thus the probability that a given particle is in compartment i at time t is $p_i(t)$, i=1,2. We may write-down the differential equations

$$\frac{dp_1(t)}{dt} = -\alpha_1 p_1(t)$$
$$\frac{dp_2(t)}{dt} = \alpha_1 p_1(t) - \alpha_2 p_2(t), \qquad (5.100)$$

with initial conditions $\underline{p}(0) = (1, 0)'$. Solving equations (5.100) yields the ps defined earlier. Note that this sort of approach based on compartmental occupancy probabilities was adopted by Faddy (1992) in a different application.

We next define the indicator variable

$$X_i(t) = \begin{cases} 1 & \textit{if ith particle is in second} \\ & \textit{compartment at time t.} \\ 0 & \textit{otherwise.} \end{cases} \qquad (5.101)$$

Thus, $Pr(X_i(t) = 1) = p_2(t)$, yielding

$$E(X_i(t)) = p_2(t) \ \text{ and } \ var(X_i(t)) = p_2(t)(1-p_2(t)) \quad i=1,\ldots,\alpha_3. \qquad (5.102)$$

For each $t \in \underline{t}$, we may regard the set $\{X_i(t)\}$ as independent Bernoulli trials. Now, the number of particles in the second compartment at time t is

$$N_2(t) = \sum_{i=1}^{\alpha_3} X_i(t) \tag{5.103}$$

$$\Rightarrow N_2(t) \sim Bin(\alpha_3, p_2(t)),$$

as derived earlier. This approach offers a much more efficient and, indeed, elegant solution to finding the distribution of $N_2(t)$. However, we may also proceed further. Choosing, without loss of generality, $t_2 > t_1$, then

$$cov(N_2(t_1), N_2(t_2)) = cov\left(\sum_i X_i(t_1), \sum_j X_j(t_2)\right).$$

Since individual particles behave independently of each other, a consequence of the linearity assumption underlying the model, we have

$$cov(X_i(t_1), X_j(t_2)) = 0 \quad \text{for } i \neq j.$$

Thus,

$$cov(N_2(t_1), N_2(t_2)) = \sum_{i=1}^{\alpha_3} cov(X_i(t_1), X_i(t_2)). \tag{5.104}$$

Now,

$$\begin{aligned} cov(X_i(t_1), X_i(t_2)) &= E(X_i(t_1)X_i(t_2)) - E(X_i(t_1))E(X_i(t_2)) \\ &= Pr(X_i(t) = 1 \cap X_i(t_2) = 1) - p_2(t_1)p_2(t_2), \end{aligned} \tag{5.105}$$

the above expression following because $X_i(t)$ is a binary random variable. We proceed as follows.

$$\begin{aligned} Pr(X_i(t_1) = 1 \cap X_i(t_2) = 1) &= Pr(X_i(t_2) = 1 \mid X_i(t_2) = 1) Pr(X_i(t_i) = 1) \\ &= Pr(\textit{ith particle doesn't leave in a time } (t_2 - t_1)) p_2(t_1), \end{aligned}$$

this last step following by the lack-of-memory property. For each particle the time in the second compartment follows a pure death process, rate α_2. Therefore,

$$cov(X_i(t_1), X_i(t_2)) = p_2(t_1)[e^{-\alpha_2 (t_2 - t_1)} - p_2(t_2)], \tag{5.106}$$

and substituting this result back into (5.104) we obtain

$$cov(N_2(t_1), N_2(t_2)) = \alpha_3 p_2(t_1)[e^{-\alpha_2 (t_2 - t_1)} - p_2(t_2)]. \tag{5.107}$$

Note that on letting $t_1, t_2 \rightarrow t$ we obtain the expected Binomial variance result.

The theoretical discussion of this section is employed in the work to follow in an investigation of the efficiency of the transform methods applied to the OCO model. We make particular

reference to the symbolic algebra package MAPLE, which proved particularly useful.

5.8. Efficiency Comparisons with Least-Squares.

We concentrate on the single *s*-value approach in this work, so that efficiency may be defined as

$$E(s;\hat{\alpha}_i) = \frac{var(\hat{\alpha}_i;OLS)}{var(\hat{\alpha}_i;Transforms)} \quad i \in (1,2,3), \tag{5.108}$$

in obvious notation. We therefore investigate efficiency as a function of the transform variable *s*. We begin with a note on calculating $var(\underline{\hat{\alpha}};OLS)$, abstracted from Seber & Wild (1989) pp 21-24.

The data are modelled as

$$N_2(t_i) = n_2(t_i;\underline{\alpha}) + \varepsilon(t_i) \quad i=1,...,n,$$

recalling equation (5.54). By Taylor-series expansion, we obtain the linear approximation

$$n_2(t_i;\underline{\alpha}) \approx n_2(t_i;\underline{\alpha}^*) + \sum_{r=1}^{p} \frac{\partial n_2(t_i;\underline{\alpha})}{\partial \alpha_r}(\alpha_r - \alpha_r^*) \quad i=1,...,n, \tag{5.109}$$

where the derivatives are evaluated at $\underline{\alpha}^*$, the true value of $\underline{\alpha}$. Note that the number of unknown parameters is denoted by *p*. Thus,

$$\underline{n}_2(\underline{\alpha}) \approx \underline{n}_2(\underline{\alpha}^*) + \nabla(\underline{n}_2)(\underline{\alpha} - \underline{\alpha}^*), \tag{5.110}$$

where

$$(\nabla)_{ij} = \frac{\partial n_2(t_i;\underline{\alpha})}{\partial \alpha_j} \quad i=1,...,n; \quad j=1,...,p,$$

$$[\underline{n}_2(\underline{\alpha})]_i = n_2(t_i;\underline{\alpha}) \quad i=1,...,n.$$

The least-squares function is, then,

$$\begin{aligned} S(\underline{\alpha}) &= ||\underline{N}_2 - \underline{n}_2(\underline{\alpha})||^2 \\ &\approx ||\underline{N}_2 - \underline{n}_2(\underline{\alpha}^*) - \nabla(\underline{\alpha} - \underline{\alpha}^*)||^2 \\ &= ||\underline{\varepsilon} - \nabla(\underline{\alpha} - \underline{\alpha}^*)||^2. \end{aligned} \tag{5.111}$$

This is minimized when

$$\underline{\alpha} - \underline{\alpha}^* = (\nabla'\nabla)^{-1}\nabla'\underline{\varepsilon}, \tag{5.112}$$

from standard theory. Therefore

$$var(\underline{\hat{\alpha}}) \approx \tau' \, var(\underline{N}_2)\tau \,, \tag{5.113}$$

where $\tau = \nabla(\nabla' \nabla)^{-1}$.

The variances for each parameter combination were obtained very efficiently using MAPLE, with results summarized below. Because we are employing only an approximate result, we focus on the largest, $n = 24$, sampling scheme. Variances are quoted correct to 4sf in the tables below.

Table 5.15. $var(\underline{\hat{\alpha}})$ for OLS, Initial Dose Known to be 100 Elements.

α_1	α_2	$var(\hat{\alpha}_1)$	$var(\hat{\alpha}_2)$
0.10	0.05	0.0001982	0.00003072
0.50	0.05	0.003732	0.00002999
0.95	0.05	0.01264	0.00003007

Table 5.16. $var(\underline{\hat{\alpha}})$ for OLS, Initial Dose Unknown.

α_1	α_2	α_3	$var(\hat{\alpha}_1)$	$var(\hat{\alpha}_2)$	$var(\hat{\alpha}_3)$
0.10	0.05	100	0.002283	0.0004064	1305
0.50	0.05	100	0.007228	0.00004204	49.33
0.95	0.05	100	0.02110	0.00003533	27.30

Using these results we may now calculate the efficiencies of the transform techniques discussed in this chapter, for both methods of transform construction. Note that MAPLE was the primary tool in this work, as for OLS, although we do not dwell on the details.

5.8.1. Quadrature

The results are depicted in Figures 5.28-30 below, for the initial dose known ($\alpha_3 = 100$). Note that these plots are marked, in order of increasing s, by *, +, *. These describe the distribution of s_{opt} from the simulation study, denoting the lower quartile, median and upper quartile respectively.

Figure 5.28. $E(s)$ for $\alpha_1 = 0.10$, $\alpha_2 = 0.05$.

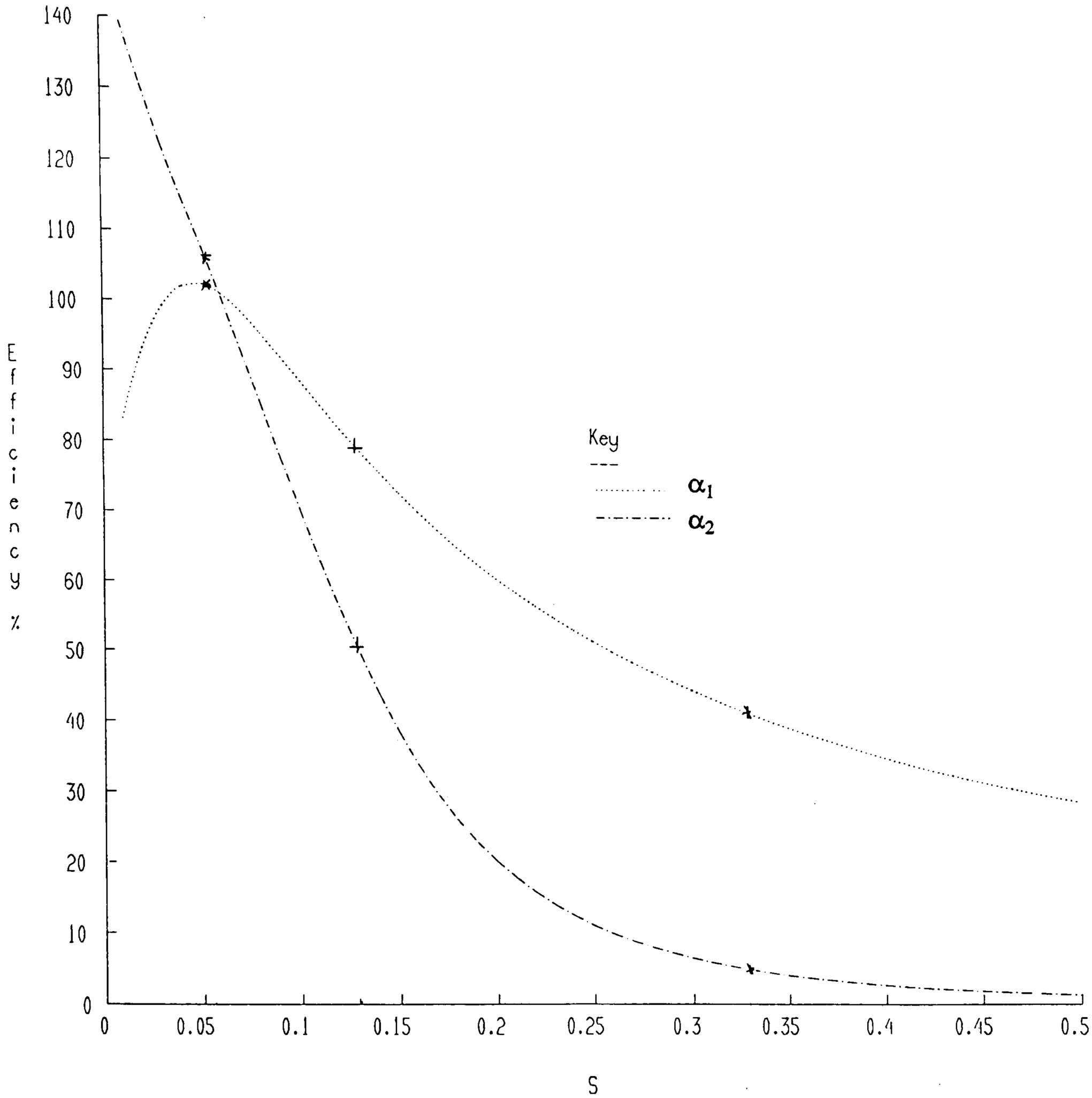


Figure 5.29. $E(s)$ for $\alpha_1 = 0.50$, $\alpha_2 = 0.05$.

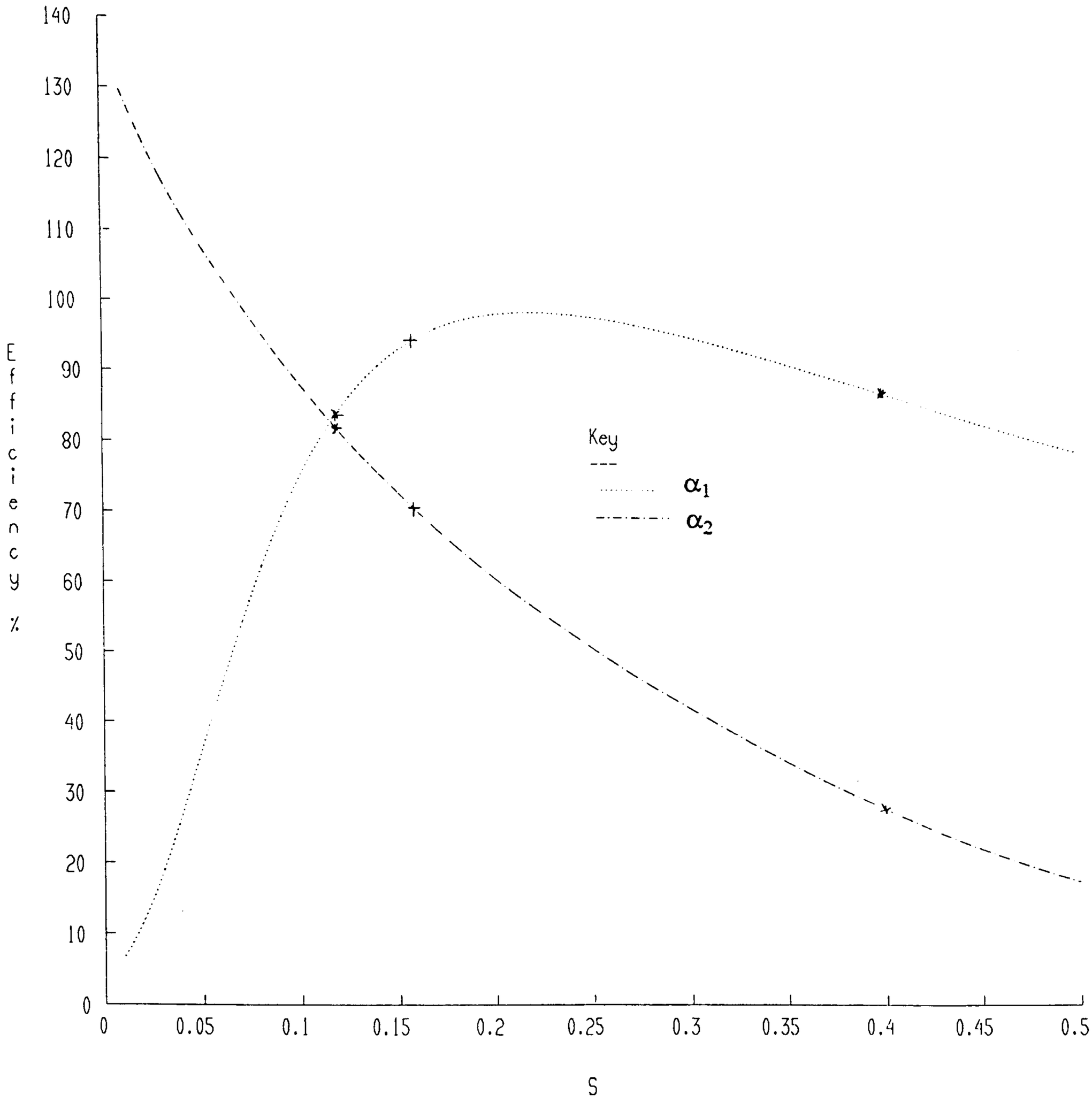
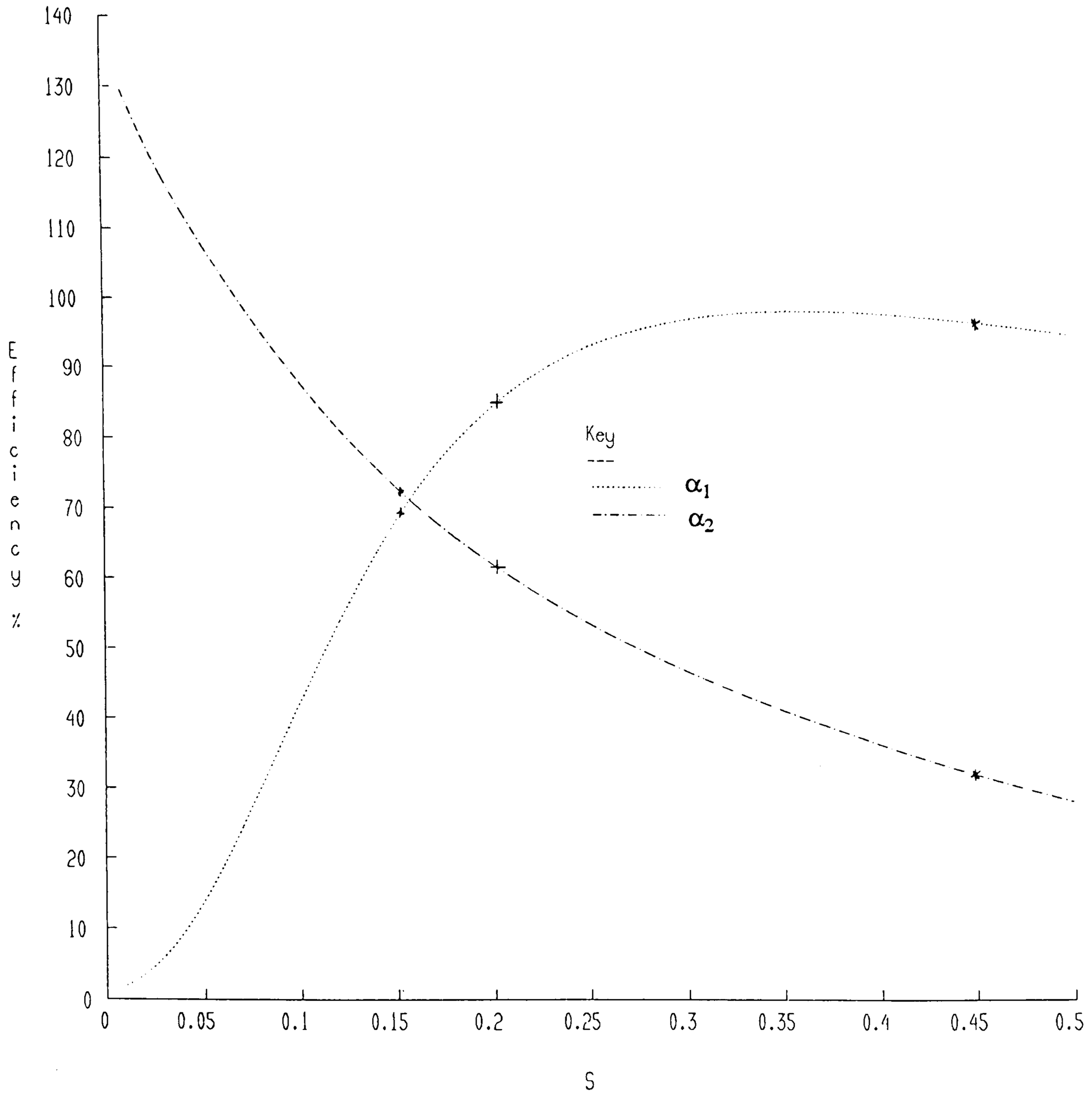


Figure 5.30. $E(s)$ for $\alpha_1 = 0.95$, $\alpha_2 = 0.05$.



At the crossing points of $E(s;\hat{\alpha}_1)$ and $E(s;\hat{\alpha}_2)$ we achieve efficiencies not less than 70%. An interesting feature is the decline in efficiency as α_1 increases, which, in the light of earlier work, we might have anticipated. The most striking feature of these plots is the divergent behaviour as $s \rightarrow 0$. The efficiency for $\hat{\alpha}_1$ tends to decline, whilst that for $\hat{\alpha}_2$ is increasing. This suggests that $\hat{\alpha}_2$ becomes strongly biased as $s \rightarrow 0$. We support this view by referring back to the work of section 5.4, where we saw that $L(0)$ tends to be poorly estimated. It is interesting to note that peak efficiency tends to occur for quite small values of s, in the region where $\hat{L}(s)$ tends to match $L(s;\underline{\alpha})$ the best.

Turning now to the distribution of s_{opt}, we see that the crossing point tends to occur at about the lower quartile in each case; there seems to be no obvious reason why this should be so. Further, it does seem surprising that such a large proportion of the optimal s should lie some distance from the peak efficiency. However, efficiency alone does not take account of bias, whereas the least-squares selection of s does, which seems to explain this phenomenon.

We consider now the case where the initial dose is unknown, with results depicted in Figures 5.31-33 below. The case $\underline{\alpha} = (0.10, 0.05, 100)'$ is very different to those discussed above. In efficiency terms, our best choice of the transform variable seems to be $s = 0$. As discussed above, this will lead to strongly biased estimators. In the other two cases we are able to find a reasonable compromise solution. Note that the divergent behaviour as $s \rightarrow 0$ has returned, whilst the decline in efficiency as α_1 increases is less pronounced. We move on now to consider the results obtained using our alternative method of transform construction.

5.8.2. Riemann-Sum

The results for α_3 known are depicted in Figures 5.34-36 below. We do not give the quartile information here, since this is essentially identical to that given earlier, as are the conclusions reached. We witness a convergence of behaviour towards the origin here, as efficiencies for both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ tend to decline. As such, the large (degenerate) efficiency peaks are not present for this technique. The results for α_3 unknown are depicted in Figure 5.37-39 at the end of this section.

Figure 5.31. $E(s)$ for $\underline{\alpha} = (0.10, 0.05, 100)'$; Quadrature-Based.

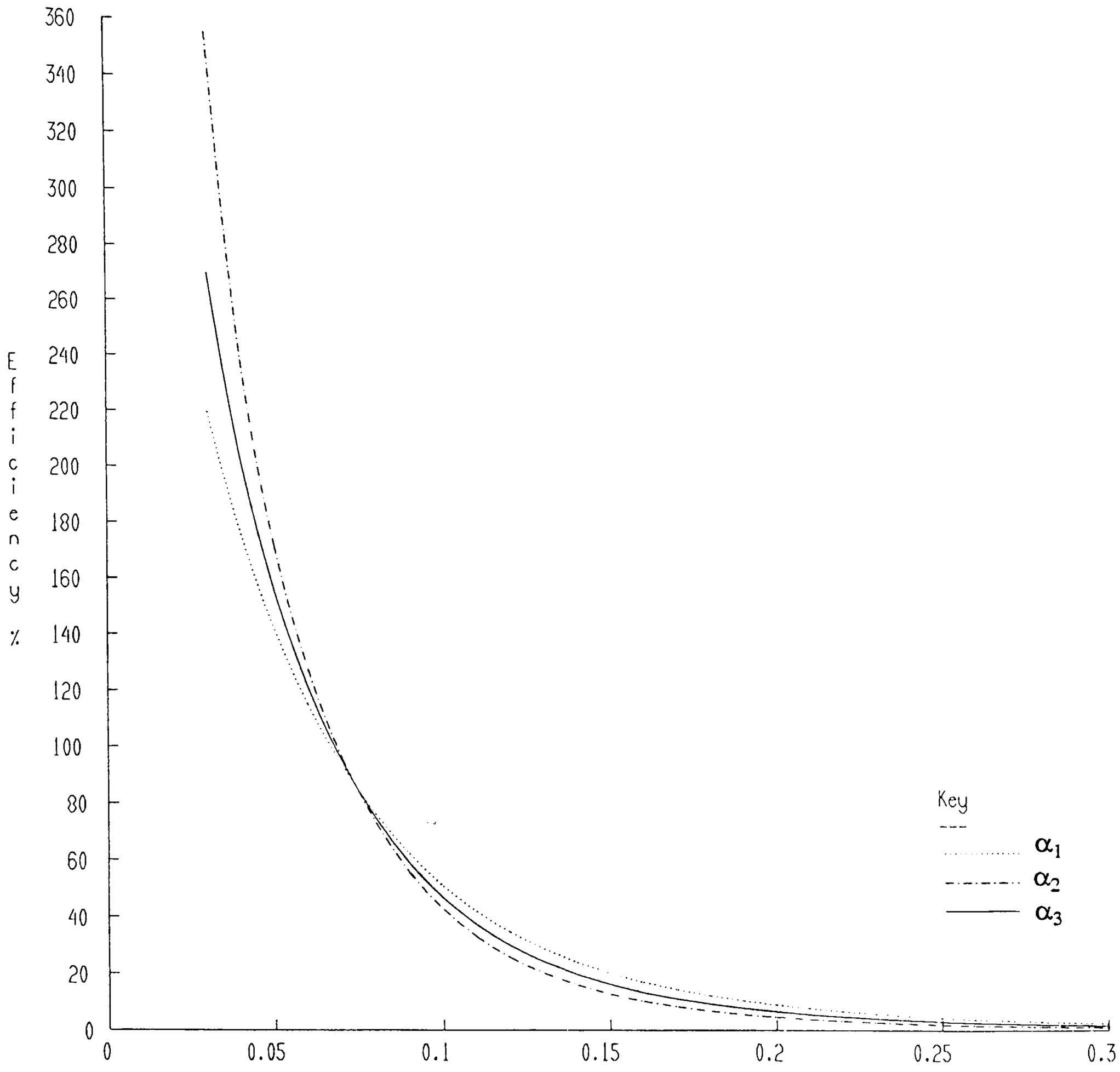


Figure 5.32. $E(s)$ for $\underline{\alpha} = (0.50, 0.05, 100)'$; Quadrature-Based.

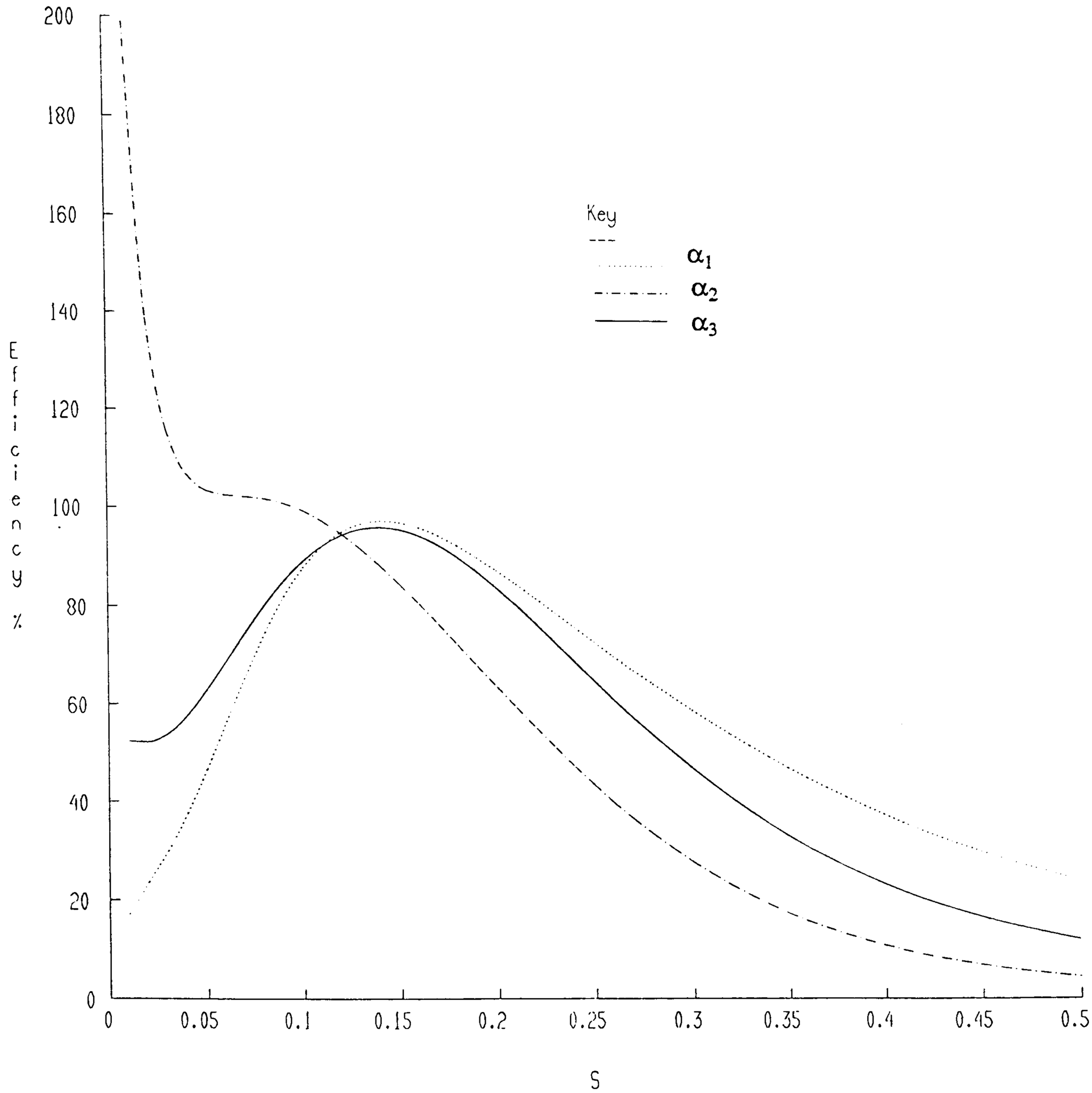
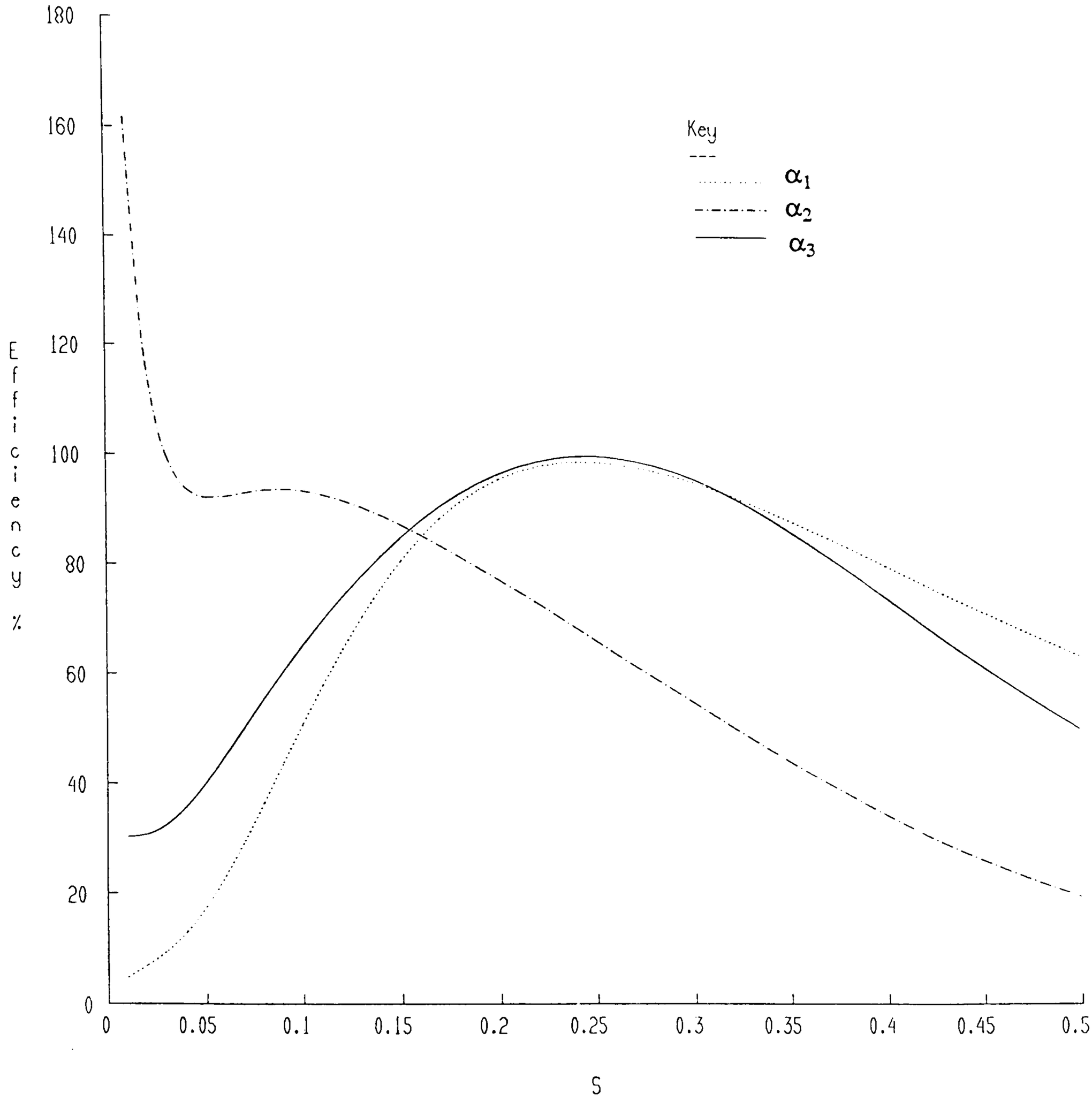


Figure 5.33. $E(s)$ for $\underline{\alpha} = (0.95, 0.05, 100)'$; Quadrature-Based.



We again witness a convergence of behaviour at the extremes of s. Overall, it seems that quadrature is capable of achieving greater efficiencies at the crossing points of the individual efficiency curves. The work of this section has established that transforms can compete well with OLS, especially when we note that the results presented here were obtained via a one-dimensional search. To find $\hat{\underline{\alpha}}_{ols}$ we must carry-out a p-dimensional search. However, p is relatively small here so that the results obtained may not generalize to larger values of p.

We conclude this section by noting that OLS can be improved upon. It arose here as a relatively simple means for selecting the transform variable. As a method for estimating $\underline{\alpha}$, we might prefer to employ weighted least-squares (WLS), exploiting our knowledge of the second-order properties of the OCO model. This was essentially the approach of Allen (1983), who employed a modified Gauss-Newton algorithm to fit the model. Seber & Wild (1989) includes a section on fitting compartment models, and discusses a number of techniques, including Allen's among them. These methods, although undoubtedly more efficient and numerically stable than OLS, are complex to apply. As such, OLS would be preferred as a means for choosing $\underline{s}$. A particular difficulty addressed by Seber & Wild is the choice of starting values, which is crucial to the success of the numerical methods discussed. It may be that transforms could play a part in supplying reliable starting values to these more complex procedures.

The results of this chapter suggest that transform methods are worthy of further consideration on grounds of numerical stability and statistical performance. Our earlier influence work suggests that they may also offer benefits in terms of robustness. We investigate this topic in the next chapter. Before proceeding to that discussion, we look now at some more general compartment models where transforms could provide a practical method of estimation.

Figure 5.34. $E(s)$ for $\alpha_1 = 0.10$, $\alpha_2 = 0.05$; Riemann-Sum Based.

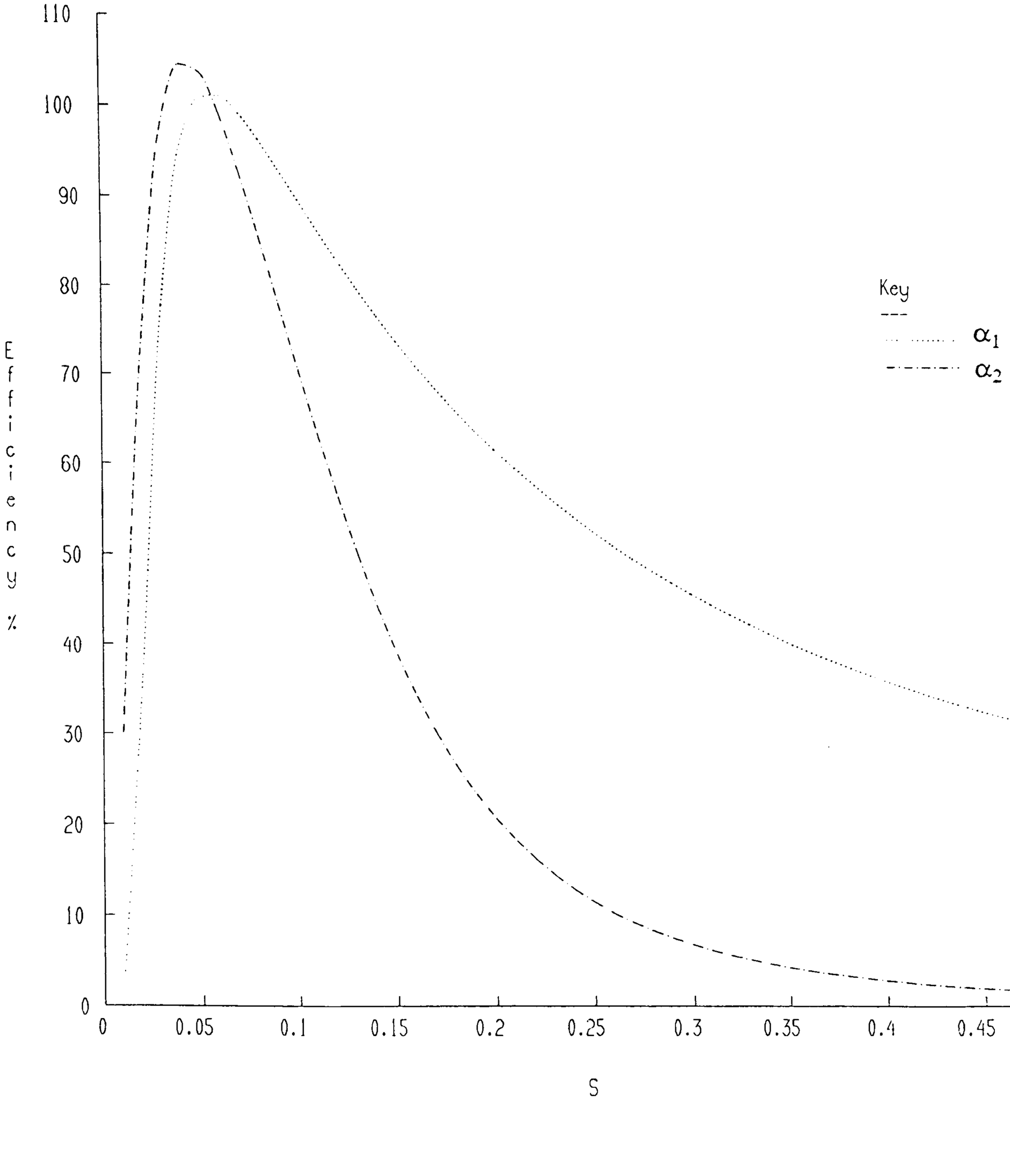


Figure 5.35. $E(s)$ for $\alpha_1 = 0.50$, $\alpha_2 = 0.05$; Riemann-Sum Based.

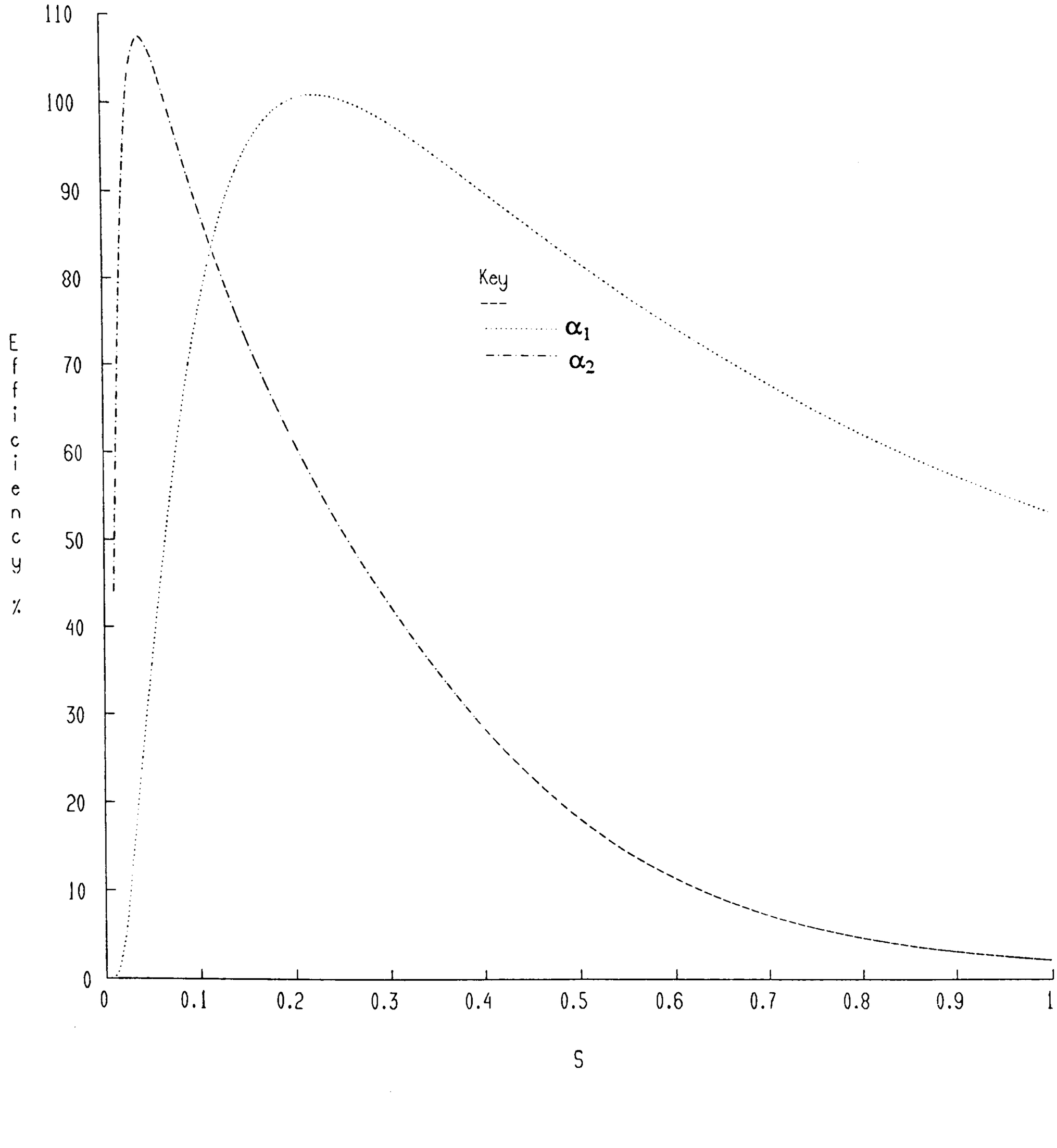


Figure 5.36. $E(s)$ for $\alpha_1 = 0.95$, $\alpha_2 = 0.05$; Riemann-Sum Based.

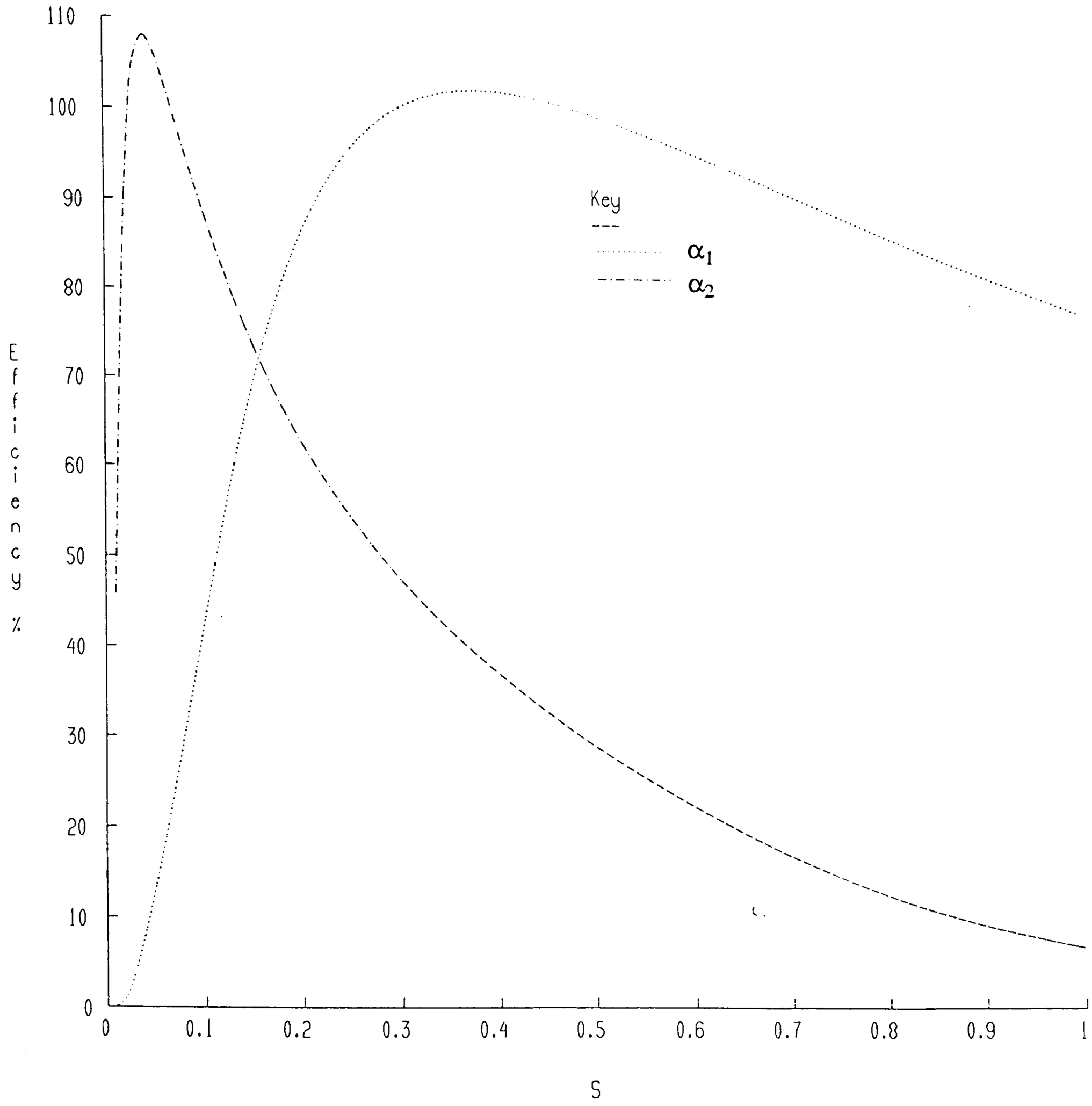


Figure 5.37. $E(s)$ for $\underline{\alpha} = (0.10, 0.05, 100)'$.

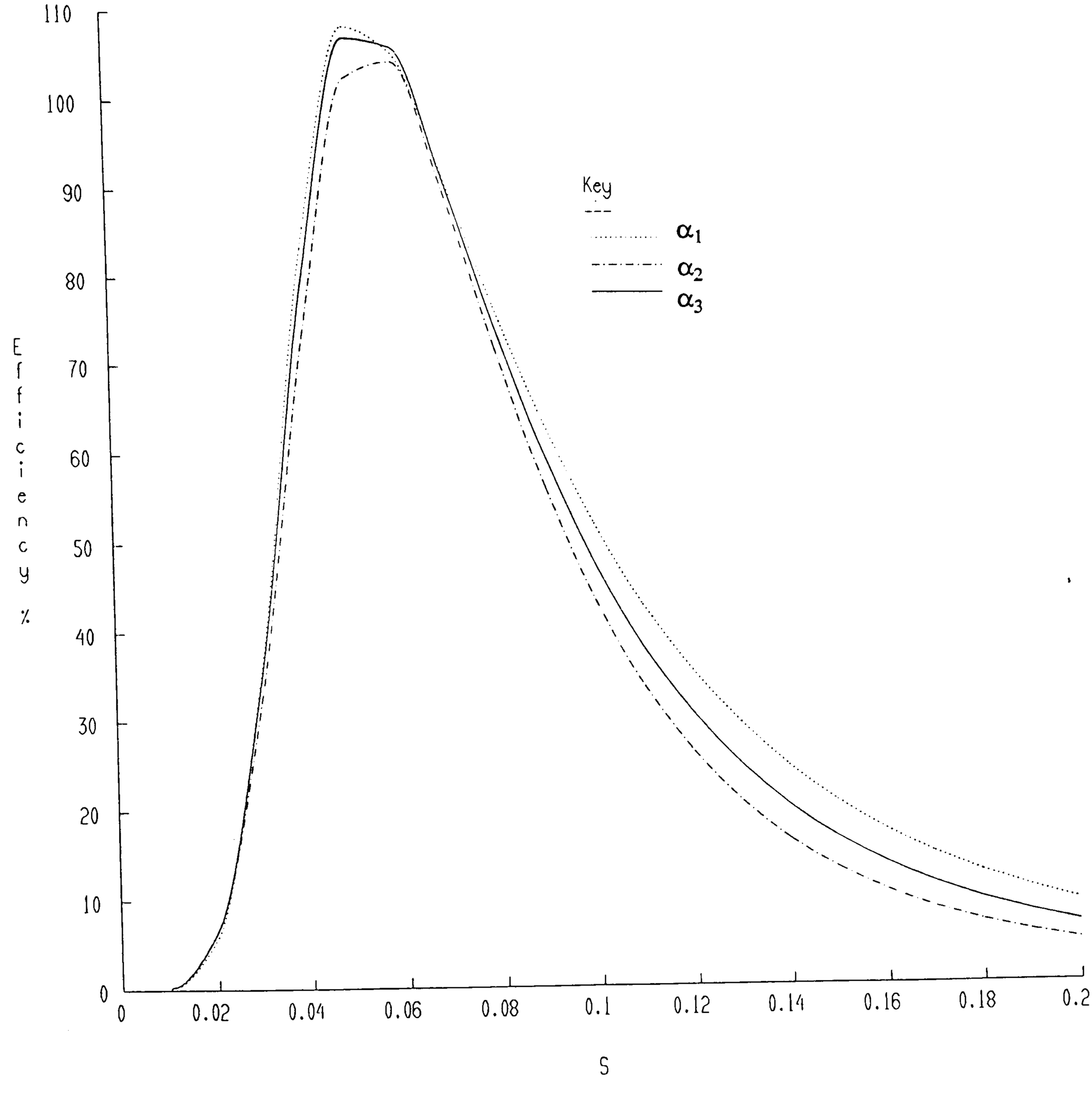


Figure 5.38. $E(s)$ for $\underline{\alpha} = (0.50, 0.05, 100)'$.

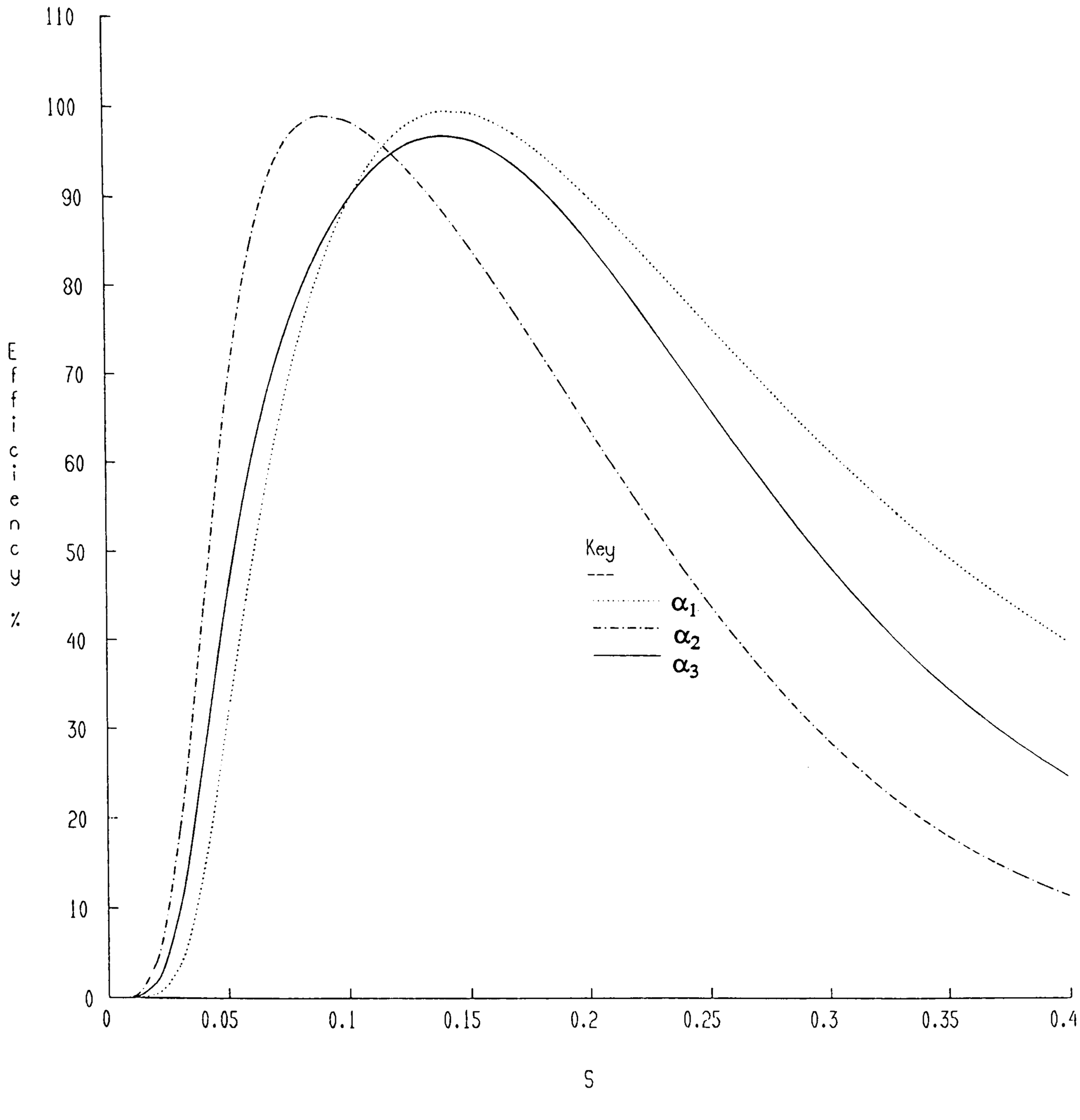
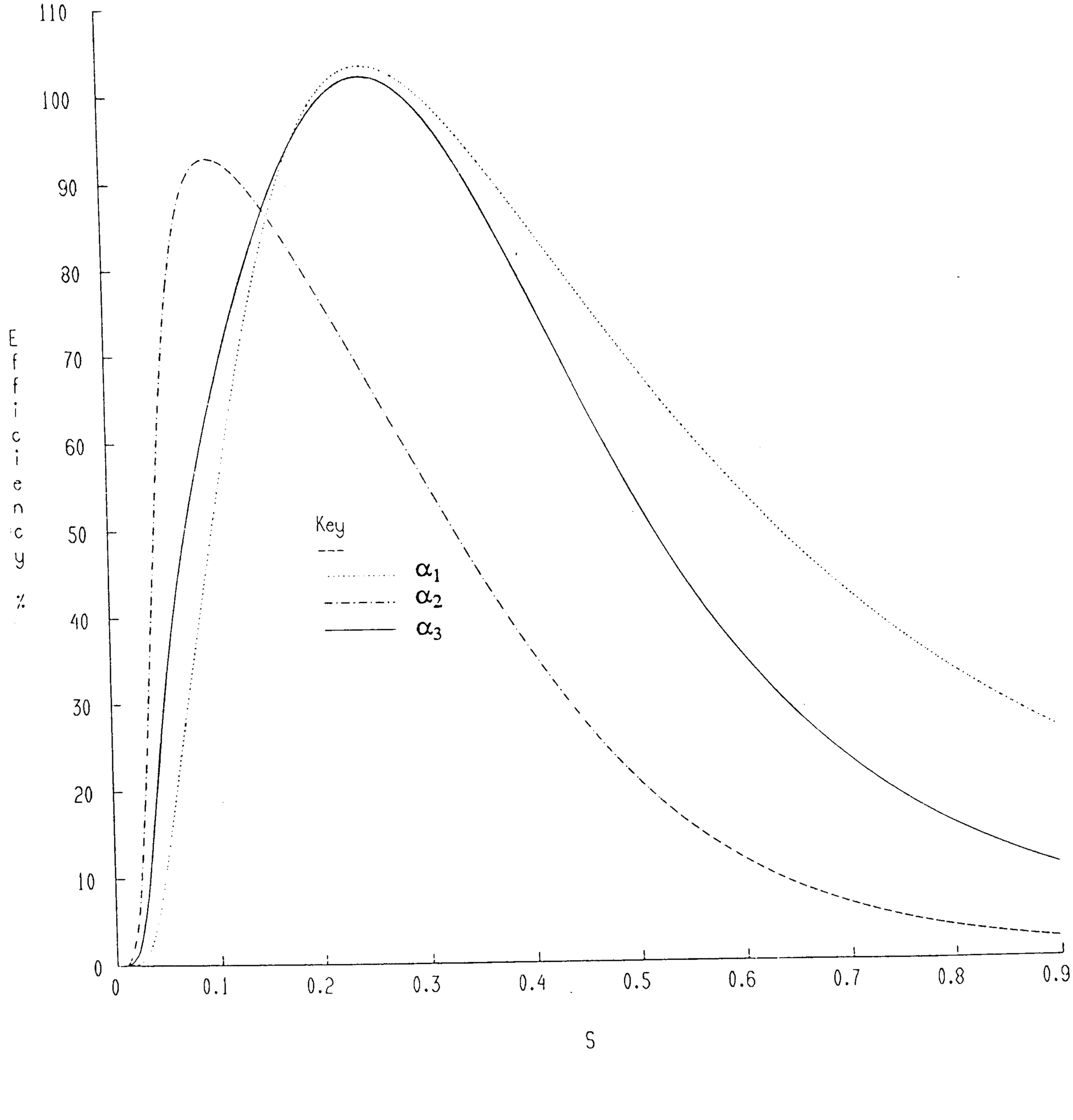


Figure 5.39. $E(s)$ for $\underline{\alpha} = (0.95, 0.05, 100)'$.



5.9. Extensions.

Many generalizations of the OCO model are possible, and we consider two alternatives here. In the first case we add more compartments to the OCO model, whilst in the second we allow for communication between compartments. Matis et al. (1989) discuss a different form of generalization. In the basic model the retention time of a particle has an exponential distribution, which they generalize to a gamma distribution.

Case 1.

It was noted by Faddy (1992) that drug retention within an organ may be modelled by a diffusion process. We can approximate this by adding more compartments, so that the process is governed by the differential equations

$$\begin{aligned} \frac{dn_1(t)}{dt} &= -\alpha_1 n_1(t) \\ \frac{dn_{k+1}(t)}{dt} &= \alpha_k n_k(t) - \alpha_{k+1} n_{k+1}(t) \qquad 1 \le k \le c-1\,, \end{aligned} \tag{5.114}$$

where the cth compartment is open. The initial condition for the process is $\underline{n}(0) = (D, 0, ..., 0)'$ in notation introduced earlier, where D represents the initial dose introduced into the first compartment at time $t = 0$.

The OCO model is the case $c = 2$, where we found

$$n_2(t) = \frac{D\alpha_1}{\alpha_1 - \alpha_2}\,(e^{-\alpha_2 t} - e^{-\alpha_1 t})$$

which may be written as

$$n_2(t) = D\alpha_1\left(\frac{e^{-\alpha_1 t}}{\alpha_2 - \alpha_1} + \frac{e^{-\alpha_2 t}}{\alpha_1 - \alpha_2}\right). \tag{5.115}$$

For the case $c = 3$ we obtain

$$n_3(t) = D\alpha_1\alpha_2\left[\frac{e^{-\alpha_1 t}}{(\alpha_2 - \alpha_1)(\alpha_3 - \alpha_1)} + \frac{e^{-\alpha_2 t}}{(\alpha_1 - \alpha_2)(\alpha_3 - \alpha_2)} + \frac{e^{-\alpha_3 t}}{(\alpha_1 - \alpha_3)(\alpha_2 - \alpha_3)}\right]. \tag{5.116}$$

These equations suggest the general form

$$n_k(t) = D\left(\prod_{j=1}^{k-1}\alpha_j\right)\sum_{j=1}^{k}\frac{e^{-\alpha_j t}}{\prod_{i\neq j}^{k}(\alpha_i - \alpha_j)}\,; \tag{5.117}$$

it remains to check whether this satisfies the system (5.114). Now,

$$\alpha_k n_k(t) - \alpha_{k+1} n_{k+1}(t) = D\left(\prod_{j=1}^{k}\alpha_j\right)\left[\sum_{j=1}^{k}\frac{e^{-\alpha_j t}}{\prod_{i\neq j}^{k}(\alpha_i-\alpha_j)} - \sum_{j=1}^{k+1}\alpha_{k+1}\frac{e^{-\alpha_j t}}{\prod_{i\neq j}^{k+1}(\alpha_i-\alpha_j)}\right].$$

We neglect the leading term $D\left(\prod_{j=1}^{k}\alpha_j\right)$ for the moment, and concentrate on the bracketed terms.

These may be written as

$$\sum_{j=1}^{k}\frac{e^{-\alpha_j t}}{\prod_{i\neq j}^{k}(\alpha_i-\alpha_j)}\left[1-\frac{\alpha_{k+1}}{\alpha_{k+1}-\alpha_j}\right] - \alpha_{k+1}\frac{e^{-\alpha_{k+1} t}}{\prod_{i\neq k+1}^{k+1}(\alpha_i-\alpha_{k+1})}$$

$$= \sum_{j=1}^{k} -\alpha_j \frac{e^{-\alpha_j t}}{\prod_{i\neq j}^{k+1}(\alpha_i-\alpha_j)} - \alpha_{k+1}\frac{e^{-\alpha_{k+1} t}}{\prod_{i\neq k+1}^{k+1}(\alpha_i-\alpha_{k+1})}$$

$$= \sum_{j=1}^{k+1} -\alpha_j \frac{e^{-\alpha_j t}}{\prod_{i\neq j}^{k+1}(\alpha_i-\alpha_j)} = \dot{n}_{k+1}(t)\,. \tag{5.118}$$

So the general form (5.117) does satisfy the system of differential equations (5.114).

If we only have data available from the open compartment, then we employ

$$L(s;\underline{\alpha},D) = \int_0^\infty e^{-st}\, n_c(t;\underline{\alpha},D)dt \tag{5.119}$$

$$= \int_0^\infty e^{-st}\, D\left(\prod_{j=1}^{c-1}\alpha_j\right)\sum_{j=1}^{c}\frac{e^{-\alpha_j t}}{\prod_{i\neq j}^{c}(\alpha_i-\alpha_j)}\,dt$$

$$\Rightarrow\ L(s;\underline{\alpha},D) = D\left(\prod_{j=1}^{c-1}\alpha_j\right)\sum_{j=1}^{c}\frac{(s+\alpha_j)^{-1}}{\prod_{i\neq j}^{c}(\alpha_i-\alpha_j)}\,. \tag{5.120}$$

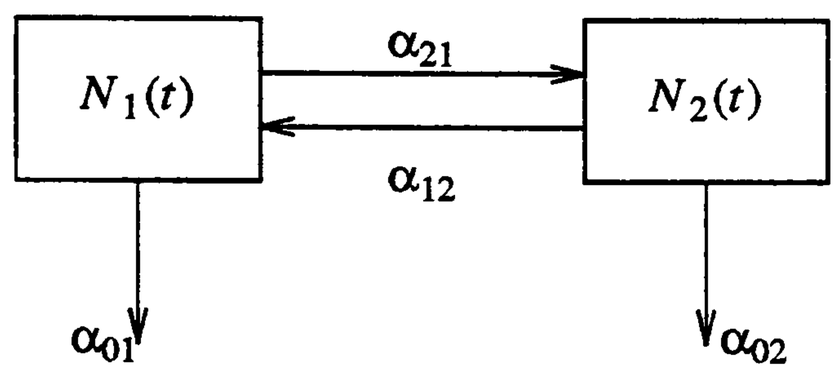
There doesn't seem to be any possibility of obtaining explicit parameter estimators in general. Indeed, it seems that we can only do this for the case $c = 2$.

Case 2.

Consider the example depicted in Figure 5.40 below, which allows for material to move in both directions. Kodell & Matis (1976) describe a weighted least-squares procedure for fitting this model. The movement of material through the system is governed by the (deterministic) differential equations

$$\begin{aligned} \dot{n}_1(t) &= \alpha_{12} n_2(t) - (\alpha_{21} + \alpha_{01}) n_1(t) \\ \dot{n}_2(t) &= \alpha_{21} n_1(t) - (\alpha_{12} + \alpha_{02}) n_2(t) \,. \end{aligned} \tag{5.121}$$

Figure 5.40. A Compartment Model Allowing for Communication.



By standard methods we find

$$\underline{n}(t) = \frac{D}{\gamma_2 - \gamma_1} \begin{bmatrix} (\alpha_{12} + \alpha_{02} - \gamma_1) e^{-\gamma_1 t} - (\alpha_{12} + \alpha_{02} - \gamma_2) e^{-\gamma_2 t} \\ \alpha_{21} (e^{-\gamma_1 t} - e^{-\gamma_2 t}) \end{bmatrix}, \tag{5.122}$$

where $\underline{n}(0) = (D, 0)'$ and $-\gamma_1, -\gamma_2$ are the roots of the characteristic equation

$$\gamma^2 + \gamma (\alpha_{21} + \alpha_{01} + \alpha_{12} + \alpha_{02}) + \alpha_{01}\alpha_{02} + \alpha_{01}\alpha_{12} + \alpha_{02}\alpha_{21} = 0 \,. \tag{5.123}$$

We present here a method due to Laurence et al. (1986). The transforms employed are

$$w_i(s) = s \int_0^\infty e^{-st} \, n_i(t) \quad i = 1,2, \tag{5.124}$$

so that

$$\begin{aligned} w_1(s) &= \frac{Ds(s + \alpha_{12} + \alpha_{02})}{(s + \gamma_1)(s + \gamma_2)} \\ w_2(s) &= \frac{Ds\alpha_{21}}{(s + \gamma_1)(s + \gamma_2)} \,. \end{aligned}$$

Assuming the initial dose to be known, then for two values s_1 and s_2 of s, we have

$$\frac{w_1(s_i)}{w_2(s_i)} = \theta(s_i) = (s_i + \alpha_{12} + \alpha_{02})/\alpha_{21} \quad i = 1,2, \tag{5.125}$$

leading to

$$\alpha_{21} = (s_1 - s_2)[\theta(s_1) - \theta(s_2)]^{-1}$$
$$\alpha_{12} + \alpha_{02} = \frac{\theta(s_2)s_1 - \theta(s_1)s_2}{\theta(s_1) - \theta(s_2)} . \tag{5.126}$$

We see also that $\alpha_{21}^{-1} = \theta'(s)$. If we now use the sum

$$w_1(s) + w_2(s) = \frac{Ds(s + \alpha_{12} + \alpha_{21} + \alpha_{02})}{(s + \gamma_1)(s + \gamma_2)}$$

and set

$$\phi(s) = Ds(s + \alpha_{12} + \alpha_{21} + \alpha_{02})(w_1(s) + w_2(s))^{-1} ,$$

then $(s + \gamma_1)(s + \gamma_2) = \phi(s)$, allowing us to deduce that

$$\gamma_1 + \gamma_2 = \frac{\phi(s_1) - \phi(s_2)}{s_1 - s_2} - s_1 - s_2 \tag{5.127}$$

and

$$\gamma_1\gamma_2 = s_1s_2 + \frac{s_1\phi(s_2) - s_2\phi(s_1)}{s_1 - s_2} . \tag{5.128}$$

As we already know α_{21} and $\alpha_{12} + \alpha_{02}$, we have

$$\alpha_{01} = (\gamma_1 + \gamma_2) - \alpha_{21} - (\alpha_{12} + \alpha_{02})$$
$$\alpha_{02} = [\gamma_1\gamma_2 - (\alpha_{12} + \alpha_{02})\alpha_{01}]/\alpha_{21} , \tag{5.129}$$

allowing us to separate $\alpha_{12} + \alpha_{02}$ to give α_{12}.

We are also able to deal with the case where the initial dose is unknown. We now take s_1, s_2, s_3 and then compute α_{21} and $(\alpha_{12} + \alpha_{02})$ by means of simple linear regression. We next evaluate

$$w_1(s) + w_2(s) = \frac{Ds(s + \alpha_{12} + \alpha_{21} + \alpha_{02})}{(s + \gamma_1)(s + \gamma_2)}$$

for the three values of s. Let

$$\kappa^{-1}(s) = \frac{w_1(s) + w_2(s)}{s(s + \alpha_{12} + \alpha_{21} + \alpha_{02})} ,$$

then we form

$$\frac{\kappa(s_i)}{\kappa(s_{i+1})} = \frac{(s_i+\gamma_1)(s_i+\gamma_2)}{(s_{i+1}+\gamma_1)(s_{i+1}+\gamma_2)} = \eta_i \quad i=1,2;$$

equations which enable us to estimate γ_1 and γ_2. Back-substitution then gives D. Laurence et al. note that these methods seemed to work well in a number of practical examples.

It seems clear from the above discussion that the method-of-moments will quickly become impractical when further compartments are added. In more general compartment models it seems that explicit parameter estimators are difficult to come by.

CHAPTER 6:

INFLUENCE THEORY FOR INDEXED RANDOM VARIABLES.

1. Introduction.

The work of Chapter 5 established that transform methods are worthy of consideration in competition with least-squares. We would anticipate from our earlier influence results that such procedures would also offer benefits in terms of robustness in the context of indexed random variables. This proposition is the subject matter for this chapter.

In section 2 a framework is introduced for calculating a measure of empirical influence, which is directly comparable to that for non-indexed random variables. Indeed, we again employ the device of empirical influence to allow easy passage to the influence function itself. We show that this technique yields the usual influence functions in simple linear regression, and subsequently that this is also true for the general linear model. It is obvious from this discussion that the technique proposed here has a great advantage of simplicity over that normally applied.

In common with the non-indexed case, there is a need to extend the fundamental definition to deal with functions and collections of statistics, which is the topic of section 3. These developments allow us, as shown in section 4, to calculate influence functions for moment estimators. Interestingly, the usual influence-theory variance result is seen to hold. The final section of Chapter 6 is a discussion of influence theory for non-linear least-squares, allowing a comparison with the transform-based estimators to be made.

2. The Influence Paradigm.

The formal definition of the influence function was motivated in Chapter 2 via the device of empirical influence, and we proceed along the same route here. For this discussion we return to a general setting, so that we collect data $\{Y(t_1),...,Y(t_n)\}$. The indexing of the random variables will be mirrored in the influence function, which is really a collection of influence functions indexed by $\{t_j : j=1,...,n\}$. We motivate the discussion to follow by consideration of the empiri-

cal transform

$$\hat{G}(s) = \sum_{k=1}^{n} h_k(s, \underline{t})Y(t_k) , \tag{6.1}$$

which estimates

$$G(s;\Theta) = \int_0^{\infty} g(s,t)\,\mu(t;\Theta)dt.$$

We consider here models of the form $Y(t) = \mu(t;\Theta) + \varepsilon(t)$ discussed in the previous chapter, so that $\mu(t;\Theta) = E(Y(t))$.

In Chapter 2 the empirical influence was found by considering the effect of an "added" observation. Under the present scenario we consider the effect on the empirical transform of adjusting the jth observation, so that (Y_j,t_j) becomes (y,t_j). Now, the empirical transform may be written as

$$\hat{G}(s) = \sum_{k \neq j}^{n} h_k(s, \underline{t})Y(t_k) + h_j(s, \underline{t})Y(t_j)$$

and so the "disturbed" transform estimator is

$$\tilde{G}(s) = \sum_{k \neq j}^{n} h_k(s, \underline{t})Y(t_k) + h_j(s, \underline{t})y . \tag{6.2}$$

A measure of empirical influence is, then,

$$\begin{aligned} EI(y,t_j;\hat{G}) &= \tilde{G}(s) - \hat{G}(s) \\ &= h_j(s, \underline{t})[y - Y(t_j)] \qquad j=1,...n. \end{aligned} \tag{6.3}$$

In order to remove dependence on a particular sample we substitute model quantities, obtaining an Influence Function:

$$IF(y,t_j;\hat{G}) = h_j(s, \underline{t})[y - \mu(t_j;\Theta)] . \tag{6.4}$$

This method is illustrated by the following

Example 6.1. We consider here the One-Compartment Open (OCO) model again, which we have modelled as

$$N_2(t) = n_2(t;\underline{\alpha}) + \varepsilon(t) ,$$

in notation established in the previous chapter. For estimation based on the Laplace transform, as

defined by equation (5.72), we obtain

$$IF(y,t_j;\hat{L}) = h_j(s,\underline{t})[y - n_2(t_j,\underline{\alpha})]\,. \tag{6.5}$$

We have two cases to consider.

(i) Gauss-Laguerre Transform Construction.

We have that $h_j = w(t_j)e^{(1-s)t_j}$, yielding

$$IF(y,t_j;\hat{L}) = w(t_j)e^{(1-s)t_j}[y - n_2(t_j;\underline{\alpha})]\,.$$

The quadrature weights, $w(.)$, have a smoothing effect, especially for larger t. We note also that the influence behaviour of the empirical Laplace transform seems to depend critically on whether $0< s < 1$ or $s \geq 1$. We noted in the efficiency work for the OCO model that efficiency seems to tail-off as s becomes large, which corresponds to reducing the sensitivity of the transform estimator. This is in tune with the trade-off between efficiency and robustness witnessed in the non-indexed work, since we see here that robustness increases with s.

In order to make further progress, we concentrate on the parameter combination $\underline{\alpha} = (0.10, 0.05, 100)'$, for which

$$IF(y,t_j;\hat{L}) = w(t_j)e^{(1-s)t_j}[\,y - 200(e^{-0.05t_j} - e^{-0.10t_j})] \qquad j=1,\ldots,n.$$

This influence surface is represented in Figure 6.1 below, for the case $n = 24$. There are two influence peaks visible. The first, towards small t, lies in a region where $N_2(t)$ should be small. As such, this peak corresponds to a large positive outlier. The second region is in the vicinity of the peak concentration and corresponds to an outlier far below the expected value. For fixed y, we see that the influence decays as t becomes large. However, for fixed t, influence increases as y increases. This is a direct result of basing estimation on the expected compartment contents.

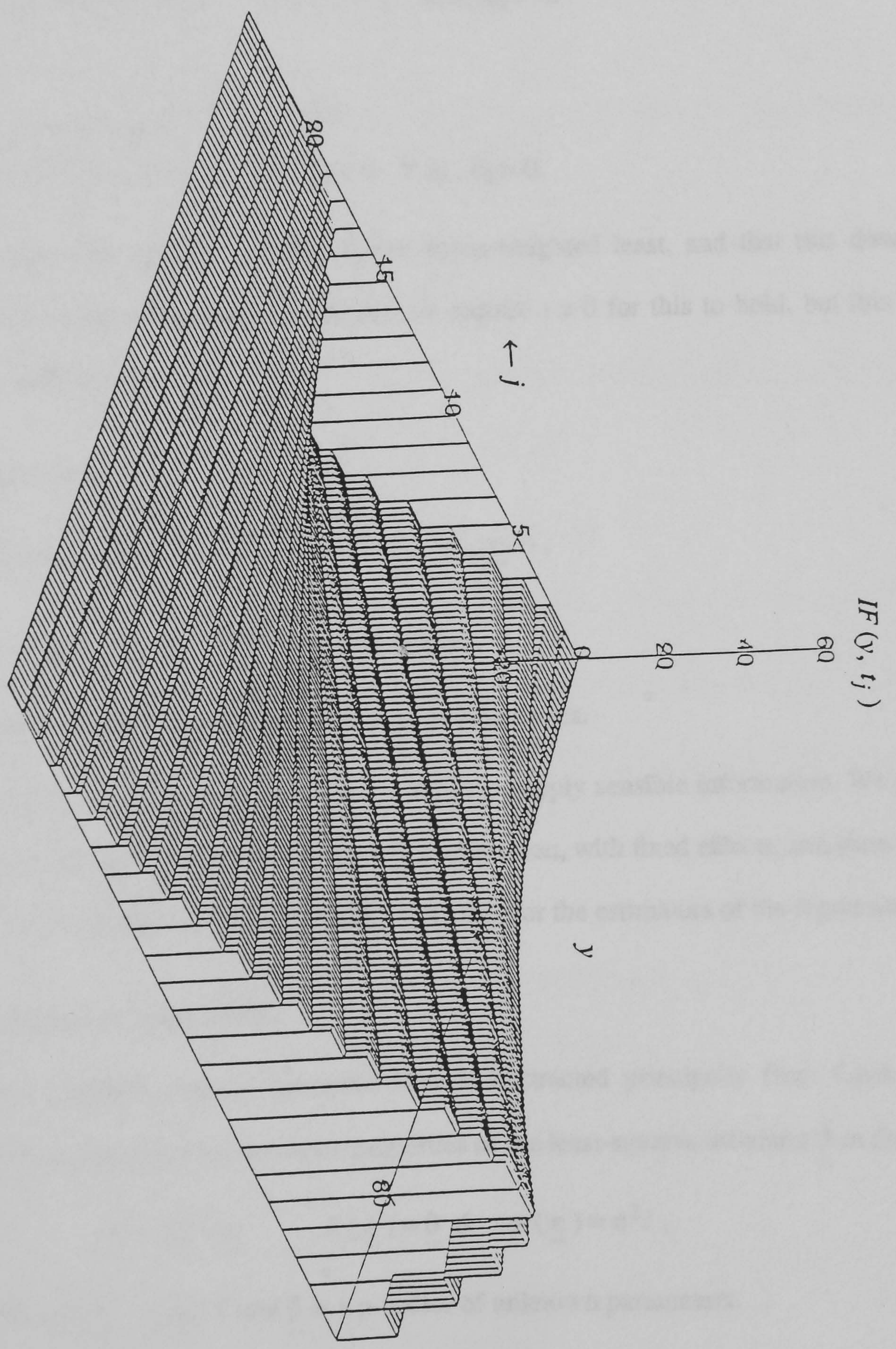
(ii) Riemann-Sum Transform Construction.

Here we have

$$h_j = s^{-1}[e^{-sc_{j-1}} - e^{-sc_j}],$$

where

Figure 6.1. Influence Surface for the OCO Model, $\underline{\alpha} = (0.10, 0.05, 100)'$, $s = 0.1$.



$$c_0 = 0,\ c_n = \infty,\ c_j = 0.5(t_j + t_{j+1}) \quad 1 \le j \le n-1\ .$$

We consider the influence functions case-by-case, as follows.

Case 1: $1 \le j \le n-1$.

$$IF(y,t_j;\hat{L}) = s^{-1}e^{-0.5st_j}[e^{-0.5st_{j-1}} - e^{-0.5st_{j+1}}](y - n_2(t_j;\underline{\alpha}))\ .$$

We may investigate influence behaviour by means of the function

$$f(x) = e^{-x}[e^{-(x-\delta_1)} - e^{-(x+\delta_2)}] \qquad x, \delta_1, \delta_2 > 0.$$

Thus,

$$f(x) = e^{-(2x-\delta_1)} - e^{-(2x+\delta_2)}$$
$$\Rightarrow\ f'(x) = 2e^{-2x}(e^{-\delta_2} - e^{\delta_1}) < 0 \quad \forall\ \delta_1,\ \delta_2 > 0.$$

We see that observations for small t_j are down-weighted least, and that this down-weighting increases monotonically with t. Note that we require $s > 0$ for this to hold, but this is in accord with our earlier work.

Cases 2,3: $j = 1$, $j = n$.

$$IF(y,t_1;\hat{L}) = s^{-1}[1 - e^{-0.5(t_j+t_{j+1})}](y - n_2(t_1;\underline{\alpha}))\ ,$$

$$IF(y,t_n;\hat{L}) = s^{-1}e^{-0.5(t_{n-1}+t_n)}(y - n_2(t_n;\underline{\alpha})).$$

We see that robustness, as before, increases as s increases.

Overall, this definition of influence seems to supply sensible information. We go on now to look at the standard approach to influence in regression, with fixed effects, and show that our very simple method yields the same influence functions for the estimators of the regression parameters.

2.1. Influence in Regression.

The standard material presented here is abstracted principally from Cook & Weisberg (1982). We investigate the influence properties of the least-squares estimator $\hat{\underline{\beta}}$ in the model

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \qquad E(\underline{\varepsilon}) = \underline{0}\ \ \&\ \ var(\underline{\varepsilon}) = \sigma^2 I\ , \tag{6.6}$$

where $X = (\underline{x}_1', \ldots, \underline{x}_n')'$ and $\underline{\beta}$ is a p-vector of unknown parameters.

The influence function of $\hat{\underline{\beta}}$ follows by considering its functional form. Letting the $(p+1)$-vector $(\underline{x}',y)$ have a joint c.d.f. F, with

$$E_F[(\underline{x}',y)'(\underline{x}',y)] = \begin{bmatrix} \Sigma(F) & \gamma(F) \\ \gamma'(F) & \tau(F) \end{bmatrix} . \tag{6.7}$$

Note that by allowing $\underline{x}$ to have design measure, we may also describe problems involving fixed effects.

The functional corresponding to the least-squares estimator of $\underline{\beta}$ is

$$T(F) = \Sigma^{-1}(F)\gamma(F) , \tag{6.8}$$

assuming that Σ is non-singular. Next, let $A(\underline{z}) = A(\underline{x},y)$ be the c.d.f. allocating all probability at $\underline{z}$. Then

$$\begin{aligned} \Sigma[(1-\varepsilon)F + \varepsilon A(\underline{z})] &= (1-\varepsilon)[\Sigma(F) + \varepsilon/(1-\varepsilon)\,\underline{x}\underline{x}'] , \\ \gamma[(1-\varepsilon)F + \varepsilon A(\underline{z})] &= (1-\varepsilon)\gamma(F) + \varepsilon y\underline{x} . \end{aligned} \tag{6.9}$$

To proceed further we make use of the general result that, given non-singularity as required,

$$(B + \underline{a}'\underline{b})^{-1} = B^{-1} - B^{-1}\underline{a}'(I + \underline{b}B^{-1}\underline{a}')^{-1}\underline{b}B^{-1} . \tag{6.10}$$

Applying this result to the first equation of (6.9), we obtain

$$[\,\Sigma + \varepsilon/(1-\varepsilon)\,\underline{x}\underline{x}'\,]^{-1} = \Sigma^{-1} - \Sigma^{-1}\varepsilon\,(1-\varepsilon)^{-1}\underline{x}\{\,I + \underline{x}'\Sigma^{-1}\varepsilon\,(1-\varepsilon)^{-1}\underline{x}\}^{-1}\underline{x}'\Sigma^{-1} . \tag{6.11}$$

These results may now be applied to the standard definition of an influence function, yielding

$$\begin{aligned} IF(y,\underline{x};\hat{\underline{\beta}}) = \lim_{\varepsilon \to 0} \{\, & \varepsilon(1-\varepsilon)^{-1}[\Sigma^{-1} - \varepsilon(1-\varepsilon)^{-1}\Sigma^{-1}\underline{x}[I + \varepsilon(1-\varepsilon)^{-1}\underline{x}'\Sigma^{-1}\underline{x}]\underline{x}'\Sigma^{-1}] \\ & [(1-\varepsilon)\gamma + \varepsilon y\underline{x}] - \varepsilon^{-1}\Sigma^{-1}\gamma \,\} . \end{aligned}$$

After a little simplification we find

$$\begin{aligned} IF(y,\underline{x};\hat{\underline{\beta}}) = \lim_{\varepsilon \to 0} \{\, & (1-\varepsilon)^{-1}\Sigma^{-1}y\underline{x} - (1-\varepsilon)^{-2}[I + \varepsilon(1-\varepsilon)^{-1}\underline{x}'\,\Sigma^{-1}\underline{x}]^{-1} \\ & \underline{x}'\,\Sigma^{-1}[(1-\varepsilon)\gamma + \varepsilon y\underline{x}] \,\} \\ = {} & \Sigma^{-1}y\underline{x} - \Sigma^{-1}\underline{x}\underline{x}'\,\Sigma^{-1}\gamma \end{aligned}$$

$$\begin{aligned} \Rightarrow \; IF(y,\underline{x};\hat{\underline{\beta}}) &= \Sigma^{-1}\underline{x}\,[y - \underline{x}'\,\Sigma^{-1}\gamma] \\ &= \Sigma^{-1}\underline{x}\,[y - \underline{x}'\,\underline{\beta}] . \end{aligned} \tag{6.12}$$

The influence for linear least-squares is, therefore, unbounded in each component as $y - \underline{x}'\underline{\beta}$ becomes large. In addition, if $\underline{x}$ is far from $E_F(\underline{x})$ and substantially in the direction of an eigenvector corresponding to a small eigenvalue of $\Sigma(F)$, then the component-wise influence can grow large even if $y - \underline{x}'\underline{\beta}$ is small.

An important question now arises: does our very straightforward influence paradigm yield the same results? We look first at simple linear regression in the following

Example 6.2.

We parameterize the model as

$$Y_i = \alpha + \beta(t_i - \bar{t}) + \varepsilon_i \qquad i = 1, \ldots, n,$$

so that the least-squares estimators are

$$\hat{\alpha} = \bar{Y}, \quad \hat{\beta} = \frac{\sum_i Y_i(t_i - \bar{t})}{\sum_i (t_i - \bar{t})^2}.$$

Now,

$$X = \begin{bmatrix} 1 & t_1 - \bar{t} \\ . & . \\ . & . \\ 1 & t_n - \bar{t} \end{bmatrix},$$

so that $\underline{t'_j} = (1, t_j - \bar{t})$. Since we are considering fixed effects only, we have

$$\Sigma = X'X = \begin{bmatrix} n & 0 \\ 0 & \sum_i (t_i - \bar{t})^2 \end{bmatrix}.$$

The influence-theoretic approach tells us that

$$IF(y, t_j; (\hat{\alpha}, \hat{\beta})') = \Sigma^{-1} \underline{t_j}(y - \underline{t_j}'\underline{\beta})$$

$$= \begin{bmatrix} n^{-1} & 0 \\ 0 & 1/\sum_i (t_i - \bar{t})^2 \end{bmatrix} \begin{bmatrix} 1 \\ t_j - \bar{t} \end{bmatrix} (y - \alpha - \beta(t_j - \bar{t})).$$

On multiplying this out we obtain

$$\begin{bmatrix} IF(y,t_j;\hat{\alpha}) \\ IF(y,t_j;\hat{\beta}) \end{bmatrix} = \begin{bmatrix} n^{-1}(y-\alpha-\beta(t_j-\bar{t})) \\ (t_j-\bar{t})(\sum_i (t_i-\bar{t})^2)^{-1}(y-\alpha-\beta(t_j-\bar{t})) \end{bmatrix}.$$

We apply now the influence paradigm suggested in this chapter, considering each parameter in turn.

Intercept.

$$\hat{\alpha} = \sum_{i=1}^{n} n^{-1} Y_i \Rightarrow h_i = n^{-1} \ \forall i \ .$$

$$\therefore \ EI(y,t_j;\hat{\alpha}) = n^{-1}(y-Y_j).$$

For an influence function we "substitute model quantities", thus replacing Y_j by its expectation to give

$$IF(y,t_j;\hat{\alpha}) = n^{-1}(y-\alpha-\beta(t_j-\bar{t})) \ ,$$

the same result as before.

Slope.

$$\hat{\beta} = \sum_{i=1}^{n} h_i Y_i \ ,$$

where

$$h_i = \frac{t_i - \bar{t}}{\sum_r (t_r - \bar{t})^2} \ .$$

Thus,

$$EI(y,t_j;\hat{\beta}) = \frac{t_j - \bar{t}}{\sum_r (t_r - \bar{t})^2} (y - Y_j) \ ,$$

and consequently we obtain the same influence function as before.

We see that our influence paradigm provides a very simple and efficient means for calculating influence functions for quantities of the form (6.1). We may, however, go further than simple linear regression. The general least-squares estimator (6.8) is a linear combination of the

observations. We may therefore extend the above example to the general linear model, as follows.

$$EI(y,\underline{x}_j;\hat{\underline{\beta}}) = \Sigma^{-1}X'(\underline{Y}^{(j)} - \underline{Y}),$$

where $\underline{Y}^{(j)}$ represents the amended data vector such that $Y_j \rightarrow y$. Now, if $\underline{xj}'$ is the jth row of X, then

$$EI(y,\underline{xj};\hat{\underline{\beta}}) = \Sigma^{-1}(\underline{x1},\ldots,\underline{xj},\ldots,\underline{xn}) \begin{bmatrix} 0 \\ . \\ . \\ y - Y_j \\ 0 \\ . \\ 0 \end{bmatrix}.$$

We therefore obtain

$$EI(y,\underline{xj};\hat{\underline{\beta}}) = \Sigma^{-1}\underline{xj}\,(y - Y_j), \tag{6.13}$$

so that

$$IF(y,\underline{xj};\hat{\underline{\beta}}) = \Sigma^{-1}\underline{xj}(y - \underline{xj}'\underline{\beta}), \tag{6.14}$$

as obtained, after a lot more work, by the influence-theoretic approach.

We can justify this equivalence as follows. Let F_i denote the c.d.f. of $Y(t_i)$ and A_j the atomic distribution such that $Pr(Y(t_j) = y) = 1$. Then

$$\begin{aligned} \hat{G}(s) &= \sum_{i=1}^{n} h_i(s,\underline{t})\int Y d\hat{F}_j \\ &= \sum_{i=1}^{n} g(\hat{F}_j), \end{aligned} \tag{6.15}$$

say. We see from this that

$$EI(y,t_j;\hat{G}) = \lim_{\varepsilon \to 0} \frac{g[(1-\varepsilon)\hat{F}_j + \varepsilon A_j] - g[\hat{F}_j]}{\varepsilon}. \tag{6.16}$$

Clearly our influence paradigm can be viewed in influence-theoretic terms.

Note further that we obtain an influence function by evaluating (6.16) at F_j rather than at $\hat{F}_j$. We therefore have the alternative definition of an influence function for quantities of the form (6.1) as

$$IF(y,t_j;\hat{G}) = \frac{\partial}{\partial\varepsilon}\{ g[(1-\varepsilon)F_j + \varepsilon A_j] \}, \tag{6.17}$$

where the derivative is evaluated at $\varepsilon = 0$.

We now have in place a basic rule for calculating influence functions. As in Chapter 3 we must extend its range of application for it to be of practical use, which is the subject of the next section.

3. Extensions to Functions and Collections of Statistics.

The development of the theory is similar to that of Chapter 2. We begin by considering influence for a differentiable function of a statistic, and then extend this to a collection of statistics.

Theorem 6.1.

Let $H(.)$ denote an infinitely differentiable function, and $\hat{T}$ some sample statistic whose empirical influence function is known. Then

$$IF(y,t_j;H(\hat{T})) = \sum_{r=1}^{\infty} \frac{[IF(y,t_j;\hat{T})]^r}{r!} \frac{\partial^r H}{\partial T^r},$$

where derivatives are evaluated at $T = \hat{T}$.

Proof

By definition,

$$EI(y,t_j;H(\hat{T})) = H(\tilde{T}) - H(\hat{T}),$$

where $\tilde{T} = \hat{T} + EI(y,t_j;\hat{T})$. Thus,

$$EI(y,t_j;H(T)) = H[T + EI(y,t_j;T)] - H(T)$$

$$= \sum_{r=1}^{\infty} \frac{[EI(y,t_j;\hat{T})]^r}{r!} \frac{\partial^r H}{\partial T^r},$$

on applying a Taylor series expansion. We may obtain, in the usual way, an expression in terms of IFs from this.

□

Corollary 6.1.

For assessing small perturbations, we have the result that

$$IF(y,t_j; H(\hat{T})) \approx IF(y,t_j; \hat{T})\frac{\partial H}{\partial T},$$

with an equivalent expression in terms of empirical influence.

At this stage it is perhaps helpful to consider the following

Example 6.3.

We find the empirical influence function at t_1 of $H(\hat{G}) = \hat{G}^2$, where $\hat{G} = h_1Y_1 + h_2Y_2$. Working from first principles,

$$\begin{aligned} EI(y,t_1; \hat{G}) &= (h_1y + h_2Y_2)^2 - (h_1Y_1 + h_2Y_2)^2 \\ &= h_1(y^2 - Y_1^2) + 2h_1h_2Y_2(y - Y_1) \\ &= h_1(y - Y_1)[h_1(y + Y_1) + 2h_2Y_2]. \end{aligned}$$

Alternatively, using Theorem 6.1,

$$\frac{\partial H}{\partial G} = 2G \qquad \frac{\partial^2 H}{\partial G^2} = 2.$$

Thus,

$$EI(y,t_j; \hat{G}^2) = h_1(y - Y_1)[2(h_1Y_1 + h_2Y_2) + \frac{h_1(y - Y_1)}{2}.2]$$

$$= h_1(y - Y_1)[h_1(y + Y_1) + 2h_2Y_2],$$

so that the equivalence is demonstrated for this example.

There is a striking difference between Theorem 6.1 and its non-indexed counterpart, Corollary 2.1. In that context we employed the definition

$$\begin{aligned} EI(X; H(\hat{T})) &= (n+1)[H(\tilde{T}) - H(\hat{T})] \\ &= (n+1)[H(\hat{T} + (n+1)^{-1}EI(X; \hat{T})) - H(\hat{T})]. \end{aligned}$$

Writing $\delta = (n+1)^{-1} EI(x; \hat{T})$, then provided

$$IF(X; \hat{T}) = \lim_{n \to \infty} EI(X; \hat{T})$$

exists, then δ is $O(n^{-1})$. Assuming that the above limit does exist, which will be the case for all but pathological examples,

$$IF(X; H(\hat{T})) = \lim_{n \to \infty} [\frac{\partial H}{\partial T} EI(X; \hat{T}) + O(n^{-1})]. \tag{6.18}$$

This is the real difference; higher-order derivatives are cancelled on taking the limit, a convenience which is not available to us in the indexed case.

We now consider what happens when we are interested in a collection of statistics.

Theorem 6.2.

Suppose that we have a vector of sample statistics, $\underline{\hat{T}} = (\hat{T}_1, \ldots, \hat{T}_p)'$, combined by a many-times differentiable function $H(.)$. Then

$$IF(y,t_j; H(\underline{\hat{T}})) \approx \sum_{i=1}^{p} IF(y,t_j; \hat{T}_i) \frac{\partial H}{\partial T_i} + \sum_{i,k=1}^{p} IF(y,t_j; \hat{T}_i) IF(y,t_j; \hat{T}_k) \frac{\partial^2 H}{\partial T_i \partial T_k} .$$

Proof

By definition,

$$EI(y,t_j; H(\underline{\hat{T}})) = H(\underline{\tilde{T}}) - H(\underline{\hat{T}}) .$$

The result then follows by considering the first few terms of a Taylor series expansion.

□

Corollary 6.2.

For assessing "local" perturbations, we have a similar result as for the non-indexed case:

$$IF(y,t_j; H(\underline{\hat{T}})) \approx \sum_{i=1}^{p} IF(y,t_j; \hat{T}_i) \frac{\partial H}{\partial T_i} .$$

We are now in a position to investigate moment estimation, where we concentrate on the Laplace transform for the purposes of this thesis.

4. Influence for Moment Estimators.

The parameter estimators arise as the solution of the estimating equations

$$\hat{L}(s_i) = L(s_i; \hat{\Theta}) \qquad i=1,...,p. \tag{6.19}$$

We have the result that

$$IF(y,t_j; \hat{L}(s_i)) \approx \sum_{k=1}^{p} IF(y,t_j; \hat{\theta}_k) \frac{\partial L(s_i)}{\partial \theta_k} + \sum_{k,l} IF(y,t_j; \hat{\theta}_k) IF(y,t_j; \hat{\theta}_l) \frac{\partial^2 L(s_i)}{\partial \theta_k \partial \theta_l}, \tag{6.20}$$

$$i=1,...,p; \quad j=1,...,n.$$

For investigating local perturbations we might employ the form

$$IF(y,t_j; \hat{L}(s_i)) \approx \sum_{k=1}^{p} IF(y,t_j; \hat{\theta}_k) \frac{\partial L(s_i)}{\partial \theta_k} \qquad i=1,...,p, \tag{6.21}$$

which may be expressed in matrix notation as

$$IF(y,t_j; \underline{\hat{L}}) \approx \nabla IF(y,t_j; \hat{\Theta}), \tag{6.22}$$

where

$$(\nabla)_{ik} = \frac{\partial L(s_i)}{\partial \theta_k} \qquad i,k = 1,...,p,$$

$$(IF(y,t_j; \hat{\Theta}))_k = IF(y,t_j; \hat{\theta}_k) \qquad k=1,...,p,$$

$$(IF(y,t_j; \underline{\hat{L}}))_i = IF(y,t_j; \hat{L}(s_i)) \qquad i=1,...,p.$$

We therefore obtain an analogous result to that for the non-indexed case, viz:

$$IF(y,t_j; \hat{\Theta}) \approx \nabla^{-1} IF(y,t_j; \underline{\hat{L}}). \tag{6.23}$$

It is interesting to note that if we apply the usual influence result to (6.23) we obtain

$$var(\hat{\Theta}) \approx \nabla^{-1} var(\underline{\hat{L}}) (\nabla^{-1})',$$

which corresponds to the result (5.83) obtained by the delta method earlier. Given that the influence functions derived here have been shown to follow the usual rule, as defined by equation (6.23), then this variance result follows by implication.

For the remainder of this section we return to the One-Compartment Open model, and investigate the robustness properties of some of the moment estimators derived in the previous chapter. We focus on the method based on one value of s, assuming that all parameters are

unknown. The empirical transforms employed may be written as $\hat{\underline{L}} = (\hat{L}, \hat{L}^{(1)}, \hat{L}^{(2)})'$, and it follows by the methods discussed in this chapter that

$$IF(y,t_j; \hat{L}^{(r)}(s)) = \frac{\partial^r h_j}{\partial s^r} (y - \mu(t_j; \underline{\alpha})) \quad r=0,1,2, \tag{6.24}$$

where the case $r = 0$ denotes no differentiation. Given these quantities we are able to form the influence vector $IF(y,t_j; \hat{\underline{L}})$. Symbolic algebra was again employed to calculate the influence functions, for the following parameter combinations and values of s depicted in Table 6.1. The values of s were selected on the basis of asymptotic efficiency, as discussed in Chapter 5, tabulated for both methods of transform construction. We again concentrate on $n = 24$, but conclusions for smaller sample sizes are directly analogous to those obtained here.

Table 6.1. Values of s Employed.

Parameters		Gauss-Laguerre	Riemann-Sum
α_1	α_2	s	s
0.10	0.05	0.08	0.06
0.50	0.05	0.13	0.13

$\alpha_3 = 100$ in all cases.

The influence functions are depicted in Figures 6.2-6.7 below for Gauss-Laguerre transform construction. Since the initial dose is of 100 elements, we investigate the range $0 \leq y \leq 100$. As an aid to the discussion to follow, it is helpful to note the time to peak concentration, t_{max}, for each case. The following table gives this quantity, along with the interval within which it falls. This is quoted because the time axes of the influence plots are labelled in terms of the time-index $j = 1,...,24$, rather than the times $\{t_j\}$ themselves.

Table 6.2. Times to Peak Concentration, $t_{max} = (\alpha_1 - \alpha_2)^{-1}\log(\alpha_1/\alpha_2)$.

α_1	α_2	t_{max}	Interval
0.10	0.05	13.86	11-12
0.50	0.05	5.17	7-8

These parameter combinations are discussed below.

(i) $\alpha_1 = 0.10$, $\alpha_2 = 0.05$.

Considering the IF surface for $\hat{\alpha}_1$ first, we observe a number of interesting features. For small t we see that large outliers tend to increase the value of $\hat{\alpha}_1$. This corresponds to particles entering the second compartment more quickly than expected. Moving towards t_{max}, large outliers tend to reduce $\hat{\alpha}_1$. If particles enter more slowly, then they will also be eliminated less quickly. The net effect is that more particles remain in the compartment than predicted by the assumed model. There is a corresponding positive peak for y less that $\mu(t_{max})$, where observations enter more quickly and so are eliminated earlier.

The final peak, for large t, is more difficult to interpret. To do this we must consider the surface for $\hat{\alpha}_2$ as well. An informal investigation of $var(\underline{\hat{\alpha}})$ revealed that $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are strongly negatively correlated. We notice in turn that the IF surfaces are essentially mirror images. However, the final peak is most interesting. The small peak for large t on $IF(y,t_j; \hat{\alpha}_1)$ is much less obvious than the negative peak of $IF(y,t_j; \hat{\alpha}_2)$. The combined effect is that particles enter more quickly, but exit much more slowly so that a greater number of particles are present for large t.

Both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are strongly correlated with $\hat{\alpha}_3$, which makes the interpretation of the influence surface of $\hat{\alpha}_3$ a little difficult. However, the central feature in the region of t_{max} is very clear. Large outliers in this region tend to increase the estimate of the initial dose, as we would expect. There is a corresponding reduction in the estimate for outliers below the peak concentration. Overall, the influence surfaces are conveying sensible information.

Figure 6.2. *IF* $(y,t_j;\ \hat{\alpha}_1\ )$ for $\alpha_1 = 0.10,\ \alpha_2 = 0.05$.

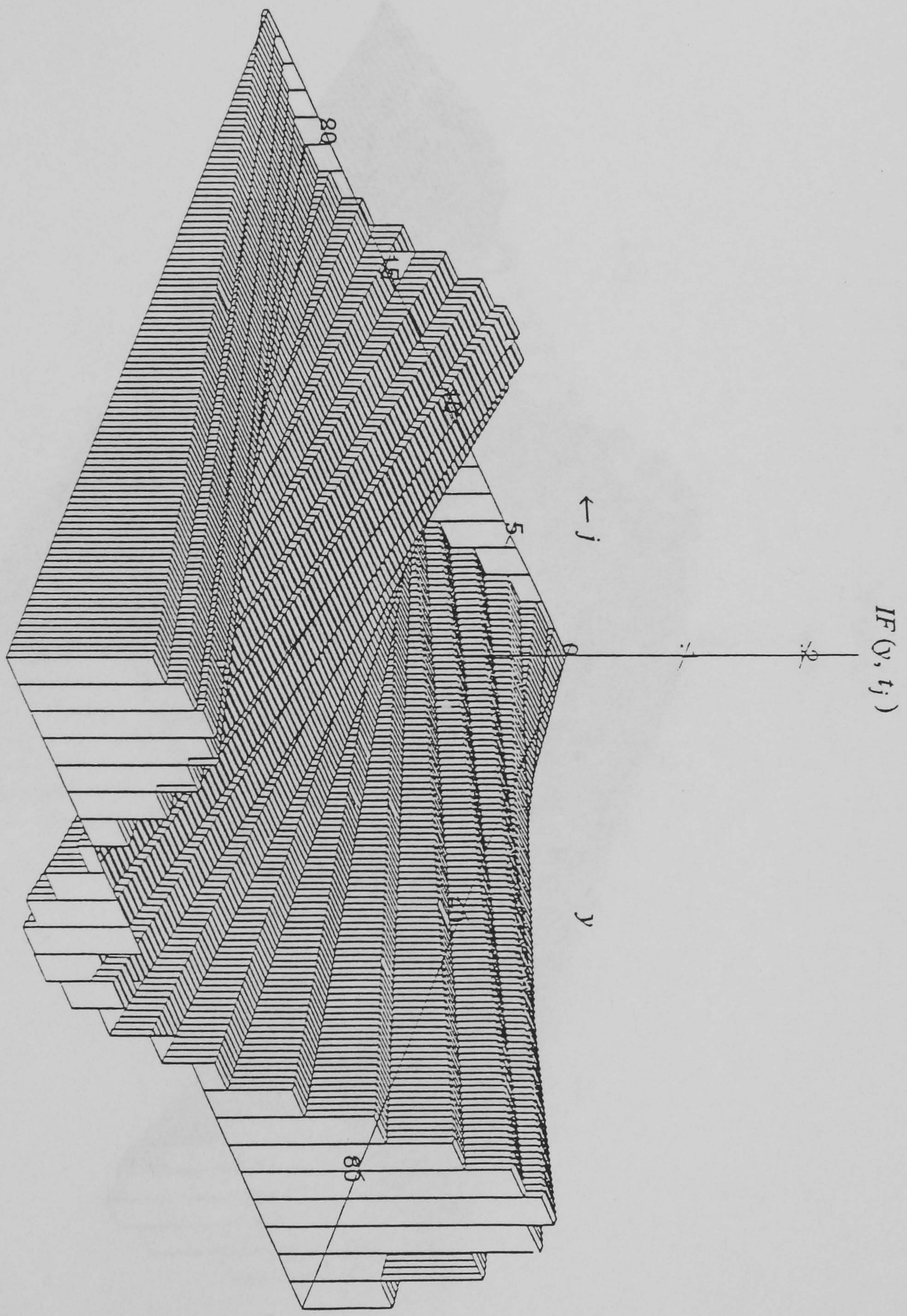


Figure 6.3. $IF\,(y,t_j;\,\hat{\alpha}_2\,)$ for $\alpha_1 = 0.10$, $\alpha_2 = 0.05$.

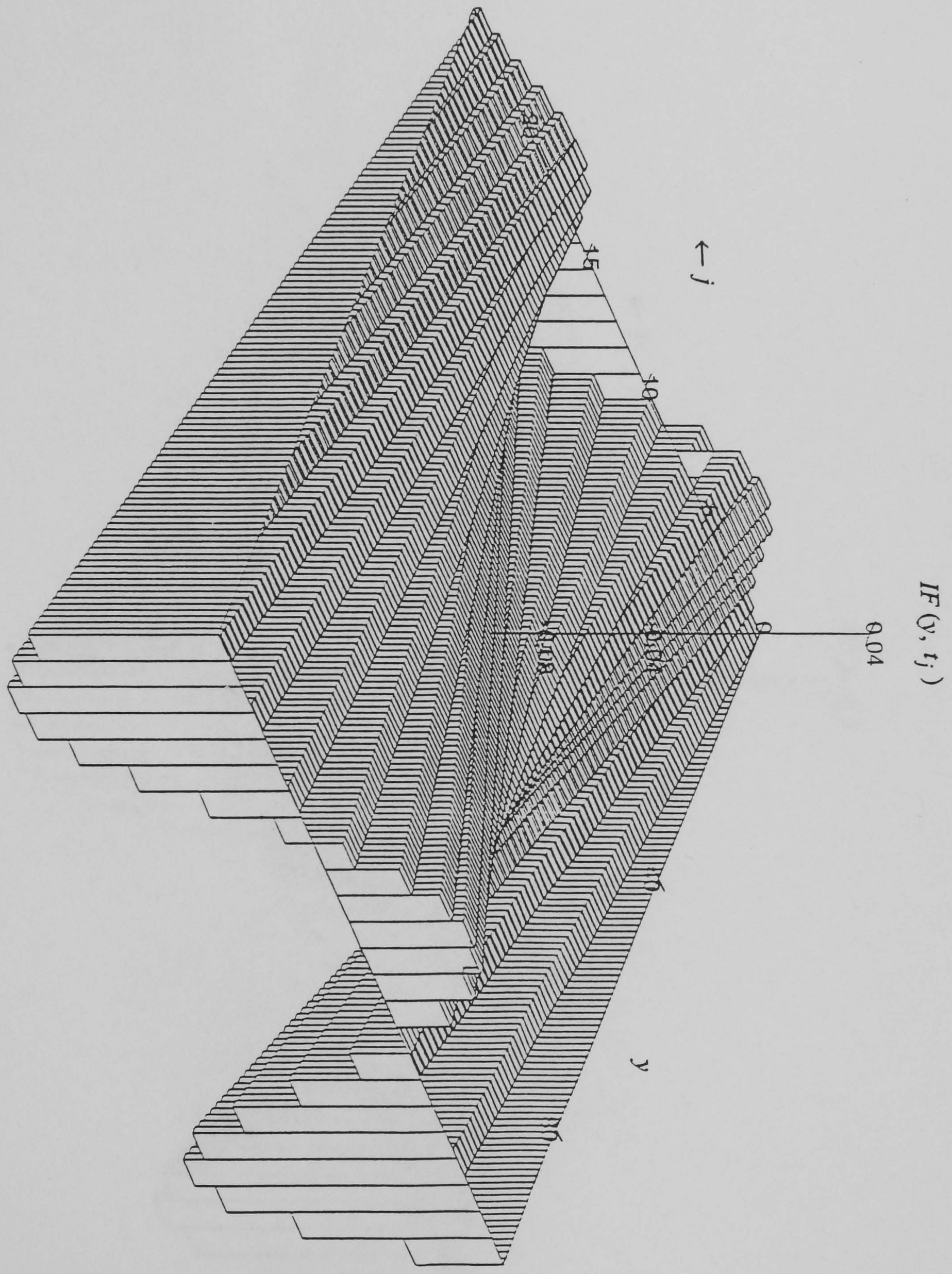


Figure 6.4. *IF* $(y,t_j;\ \hat{\alpha}_3\ )$ for $\alpha_1 = 0.10,\ \alpha_2 = 0.05$.

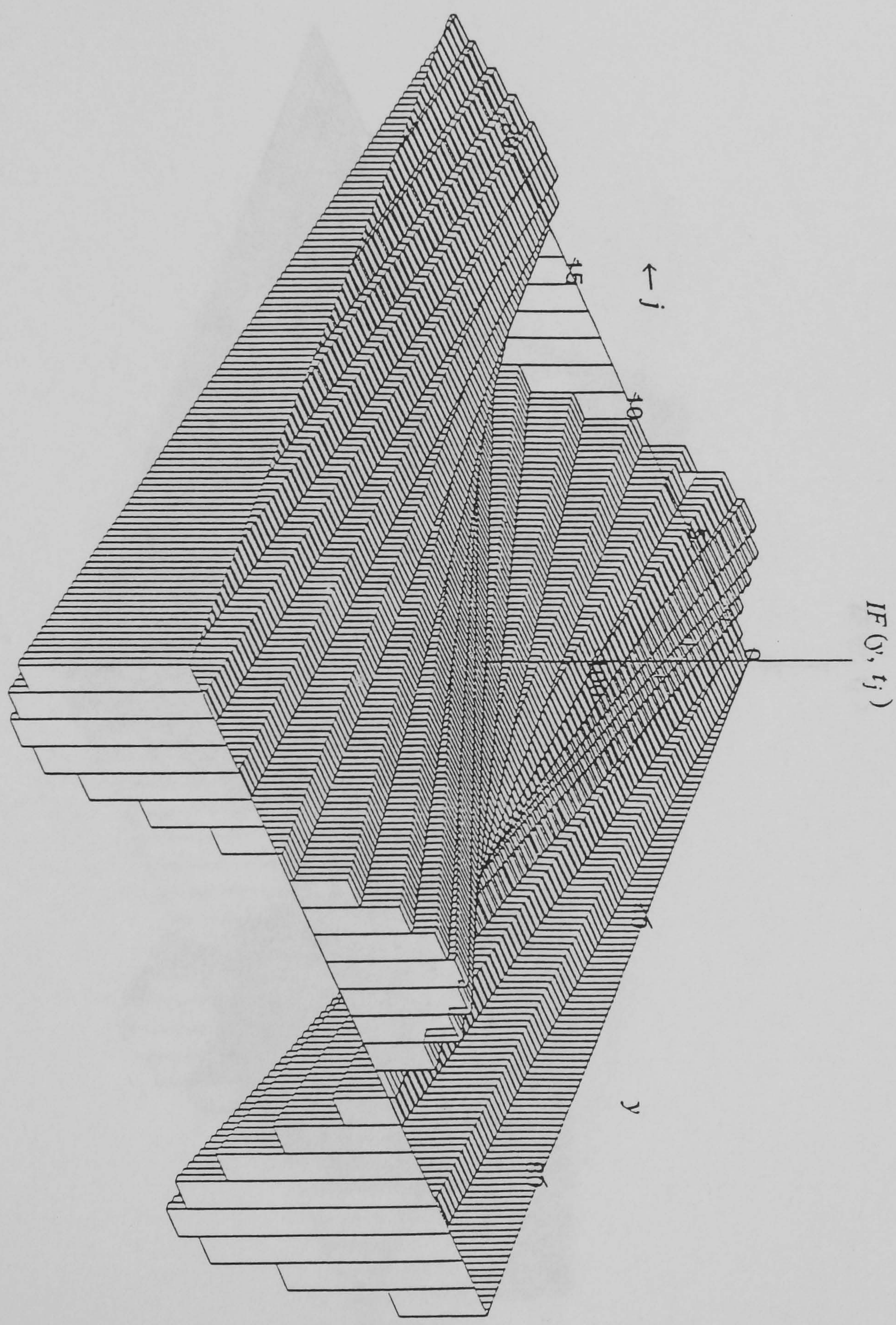


Figure 6.5. *IF* $(y,t_j;\ \hat{\alpha}_1\ )$ for $\alpha_1 = 0.50,\ \alpha_2 = 0.05$.

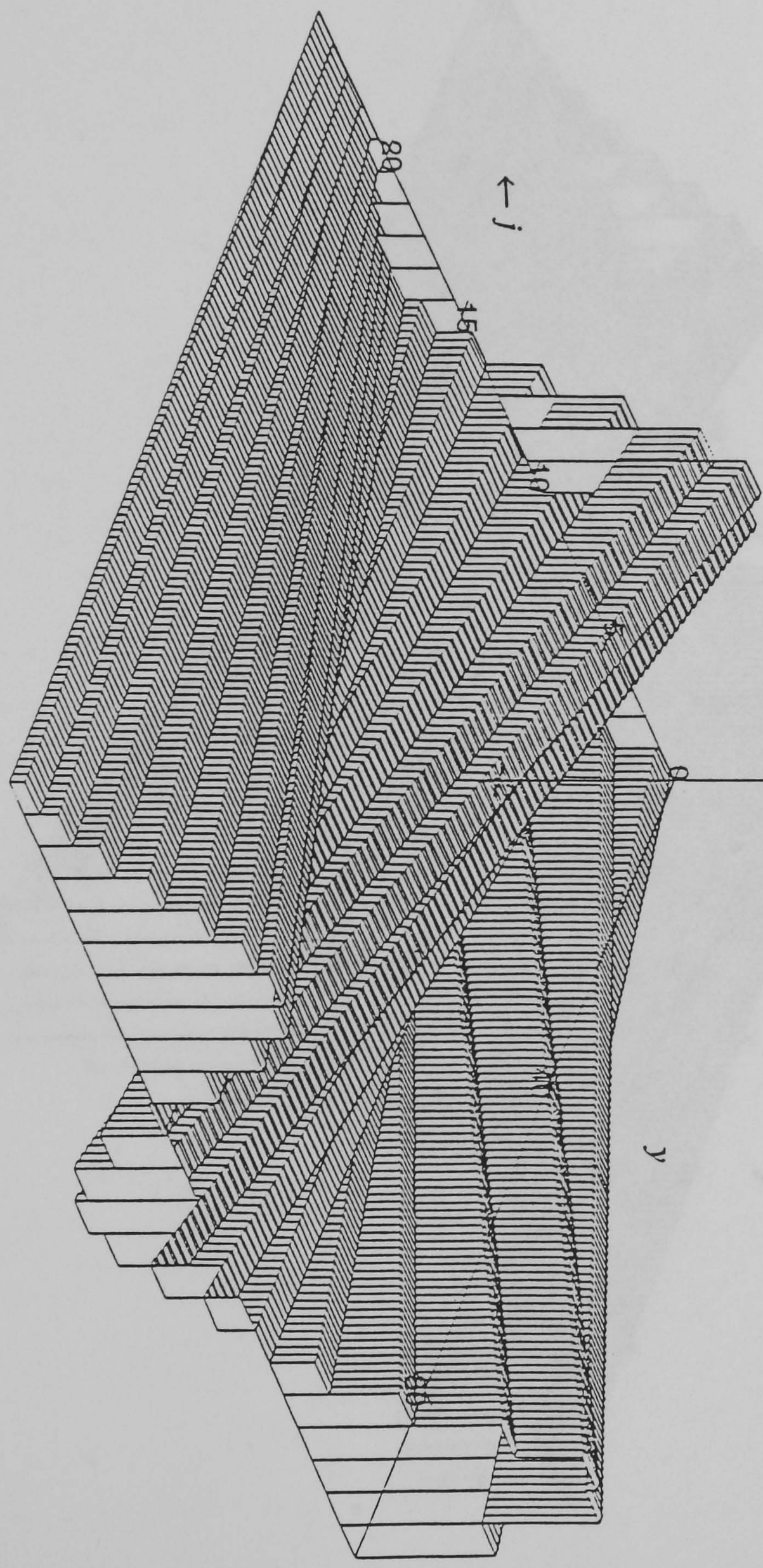


Figure 6.6. $IF\ (y,t_j;\ \hat{\alpha}_2\ )$ for $\alpha_1 = 0.50,\ \alpha_2 = 0.05$.

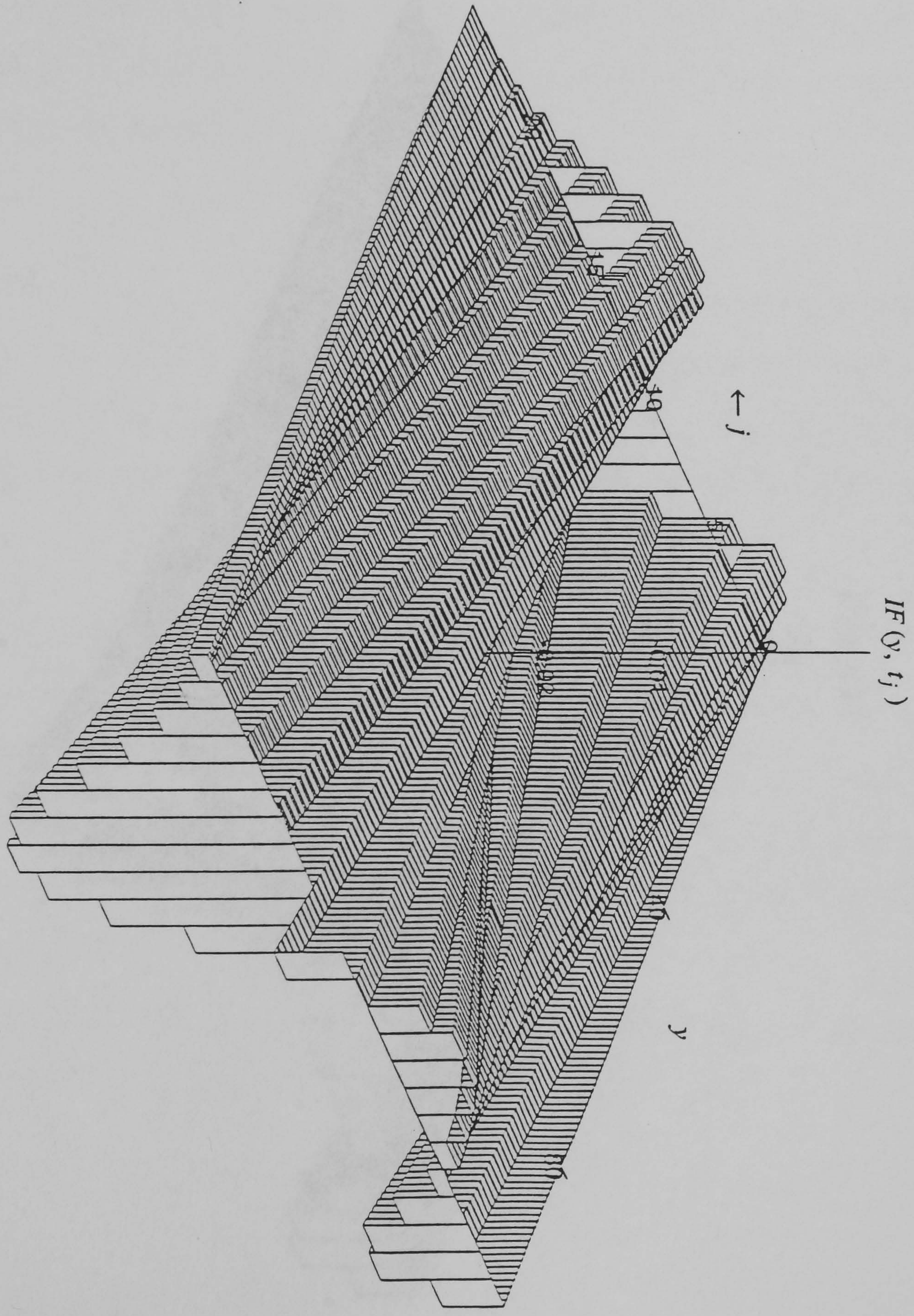
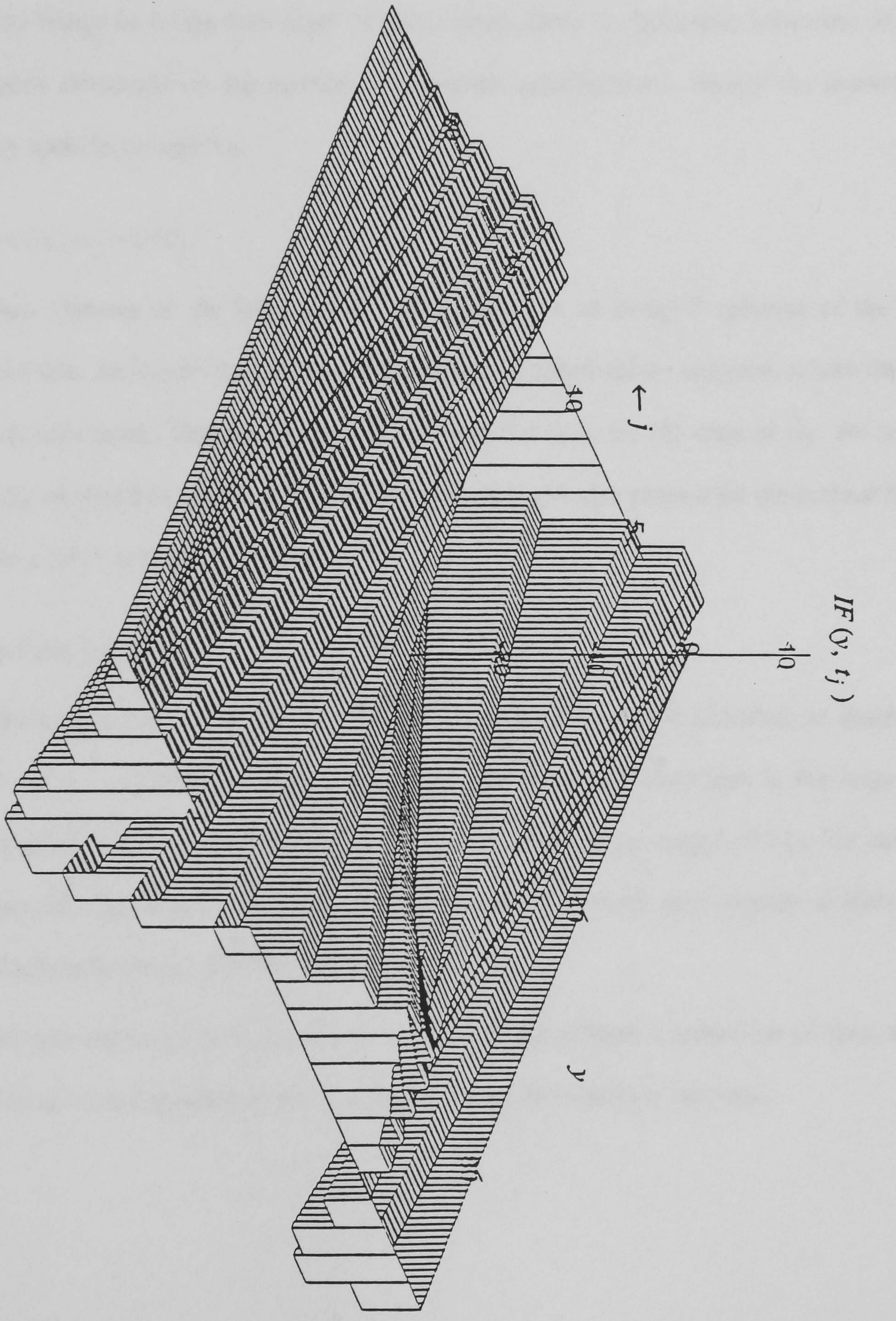


Figure 6.7. $IF\,(y,t_j;\hat{\alpha}_3)$ for $\alpha_1 = 0.50$, $\alpha_2 = 0.05$.



(ii) $\alpha_1 = 0.50,\ \alpha_2 = 0.05.$

We see, in Figures 6.5-6.7, that the same themes emerge. The only real difference is that features occur earlier in time, as should be the case given the higher transition intensity. Another feature, which is more obvious in these latter figures, is the smoothing inherent in the quadrature-based technique as t becomes large.

This brings us to the final topic of this section. Does the influence behaviour of the estimators depend obviously on the method of transform construction? Mostly the answer is no, but with very specific exceptions.

(i) $\alpha_1 = 0.10,\ \alpha_2 = 0.05.$

Most features of the influence surfaces appear to be damped versions of the the Gauss-Laguerre plots. However, the last sampling point is a spectacular exception, where there is a very large influence peak. This is depicted in Figure 6.8 below for the case of $\hat{\alpha}_1$. As before, $\hat{\alpha}_2$ is essentially obtained as a mirror image. The spike at t_{24} is also present on the surface for $\hat{\alpha}_3$, manifested as a large negative influence.

(ii) $\alpha_1 = 0.50,\ \alpha_2 = 0.05.$

For this parameter combination the influence surfaces appear identical, as demonstrated by the plot for $\hat{\alpha}_1$ in Figure 6.9 below. Note that the value of s used here is the same as that for Gauss-Laguerre, so that the transform estimators are both estimating $L(0.13)$. For case (i) above the values of s are marginally different (see Table 6.1), which may explain at least part of the difference in behaviour noted above.

We proceed in the next section to investigate the influence behaviour of least-squares, and we will then be in a position to make comparisons with transform methods.

Figure 6.8. $IF\,(y,t_j;\,\hat{\alpha}_1\,)$ for $\alpha_1 = 0.10,\ \alpha_2 = 0.05$ Based on Riemann-Sum Transform Construction.

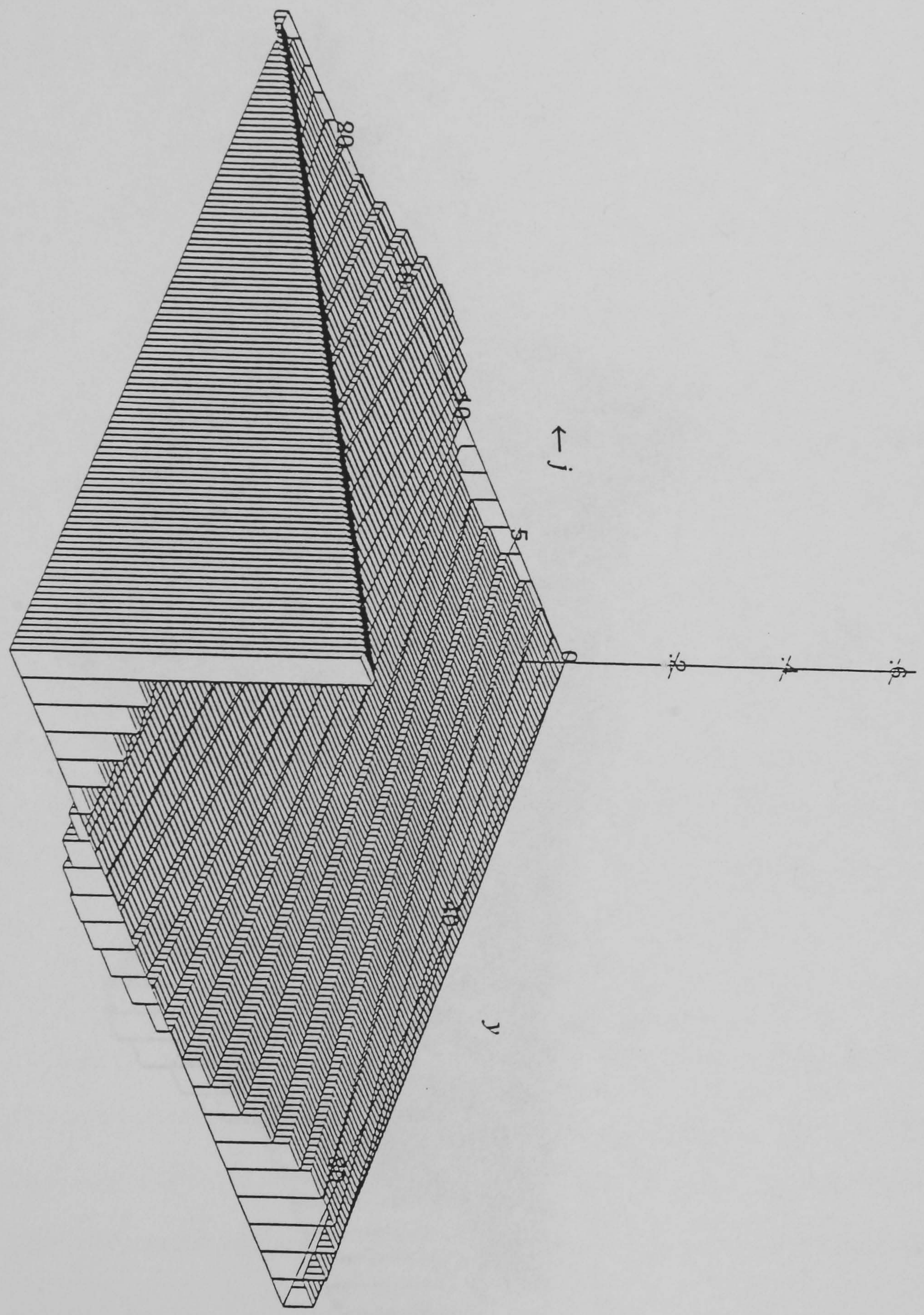
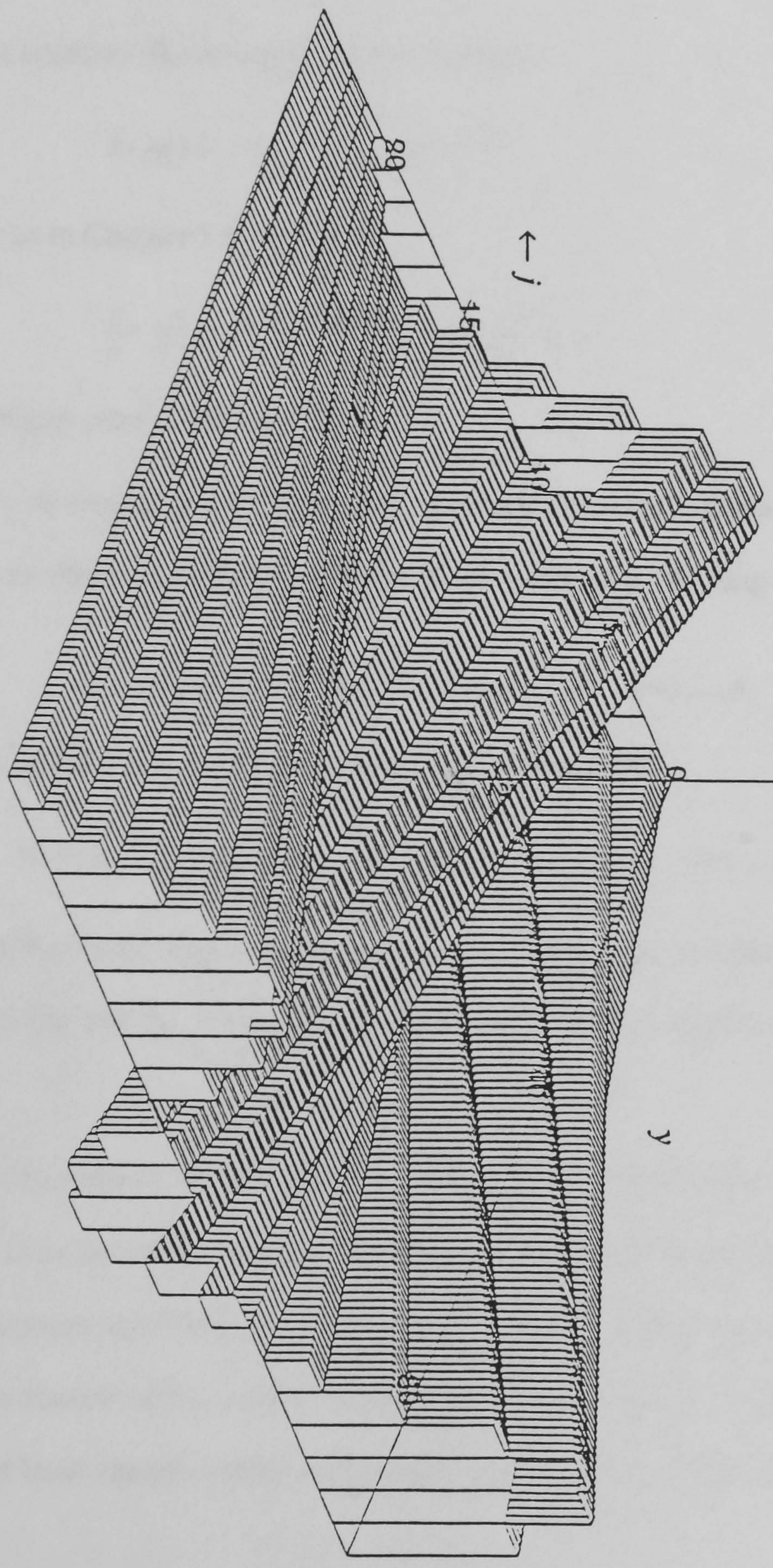


Figure 6.9. $IF\,(y,t_j;\,\hat{\alpha}_1\,)$ for $\alpha_1 = 0.50,\ \alpha_2 = 0.05$ Based on Riemann-Sum Transform Construction.



5. Comparison With Non-Linear Least-Squares.

Recall that the OCO model is expressed in this thesis as

$$N_2(t_i) = n_2(t_i; \underline{\alpha}^*) + \varepsilon (t_i) \quad i=1,...,n,$$

where $\underline{\alpha}^*$ denotes the vector of true parameters. The least-squares criterion requires minimization of

$$S(\alpha) = || N_2 - n_2(\alpha^*) ||^2 ,$$

in obvious notation. Recalling (5.111), we obtain

$$S(\underline{\alpha}) \approx || \underline{\varepsilon} - \nabla(\underline{\alpha} - \underline{\alpha}^*) ||^2$$

which led us in Chapter 5 to the result

$$\underline{\hat{\alpha}} - \underline{\alpha}^* \approx (\nabla'\nabla)^{-1}\nabla'[\underline{N}_2 - \underline{n}_2(\underline{\alpha}^*)] , \tag{6.25}$$

on appealing to standard theory.

We now exploit this explicit form to obtain influence functions to the same degree of accuracy as those obtained earlier for the transform estimators. Writing $\kappa = (\nabla'\nabla)^{-1}\nabla'$, we have

$$\hat{\alpha}_k - \alpha_k^* \approx \sum_{i=1}^{n} (\kappa)_{ki}[N_2(t_i) - n_2(t_i; \underline{\alpha}^*)] \quad k=1,...,p. \tag{6.26}$$

Applying our influence paradigm to this expression yields

$$IF(y,t_j; \hat{\alpha}_k) = (\kappa)_{kj}[y - n_2(t_j; \underline{\alpha}^*)] \quad k=1,..,p; \quad j=1,...n. \tag{6.27}$$

We see, as for transforms, that large outliers can have an unbounded impact on the estimators. The reason for this is, once again, because estimation is based on the conditional expectation $n_2(t; \underline{\alpha})$.

For the moment we consider the influence functions (6.27) in general terms for the three-parameter case, focusing on $n = 24$. These are depicted in Figures 6.10-6.12 below for the parameter combination $\alpha_1 = 0.10$, $\alpha_2 = 0.05$, $\alpha_3 = 100$. It is interesting to discuss these influence surfaces in the context of the outlier study of Rodda et al. (1975). In their work they investigated the response of least-squares versus a more robust alternative to particular patterns of outliers. They

Figure 6.10. Influence Surface for Least-squares Estimator of α_1.

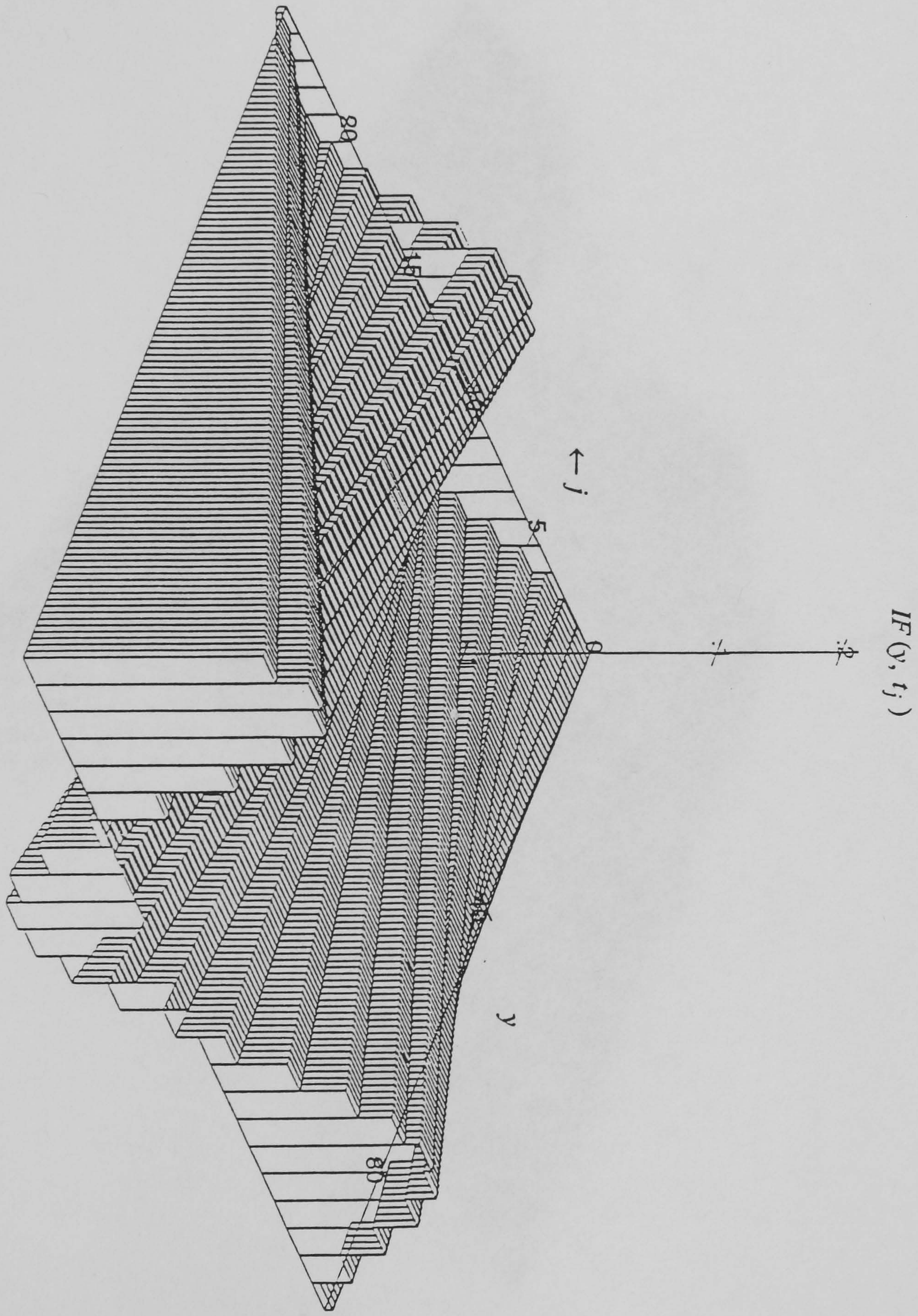


Figure 6.11. Influence Surface for Least-Squares Estimator of α_2.

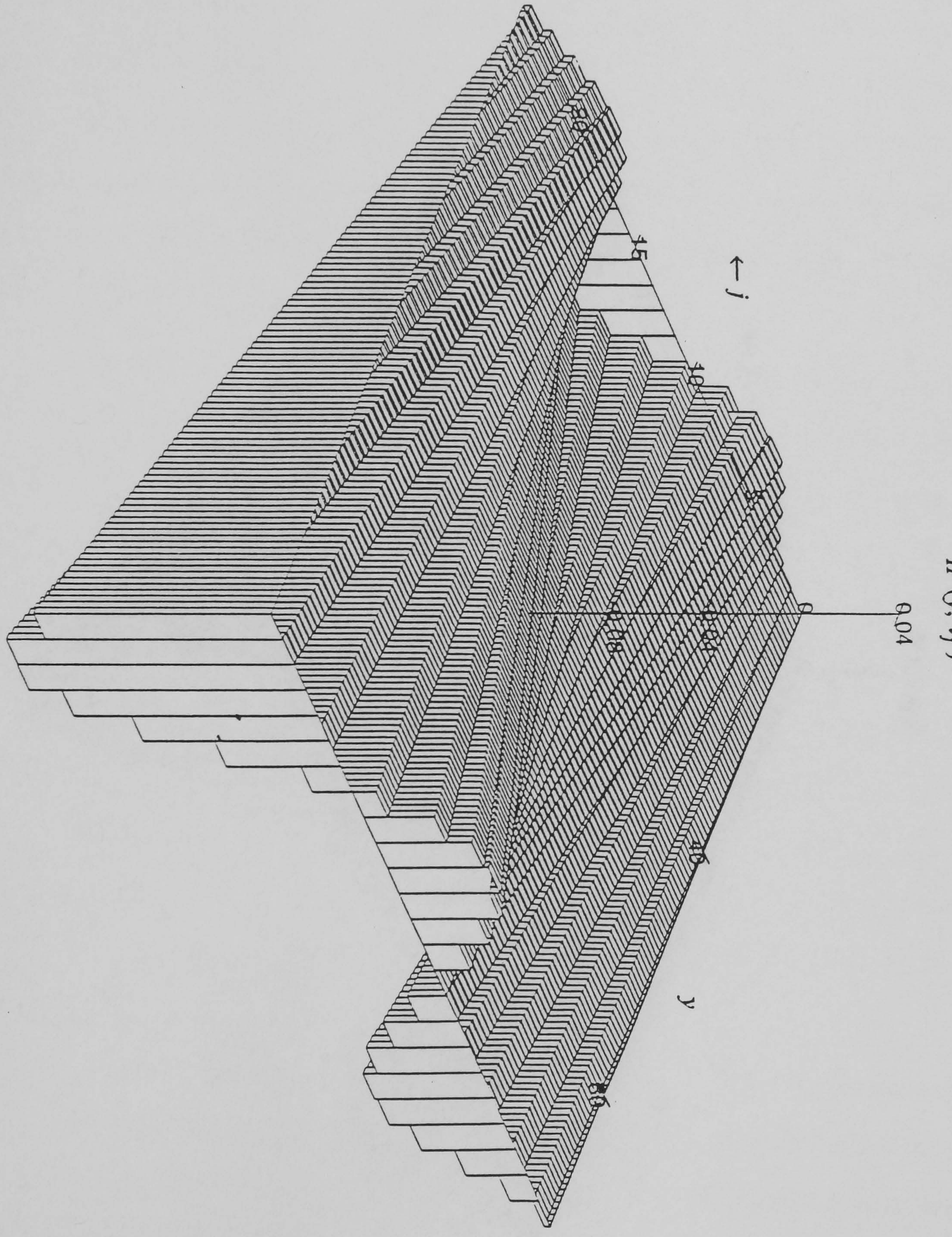
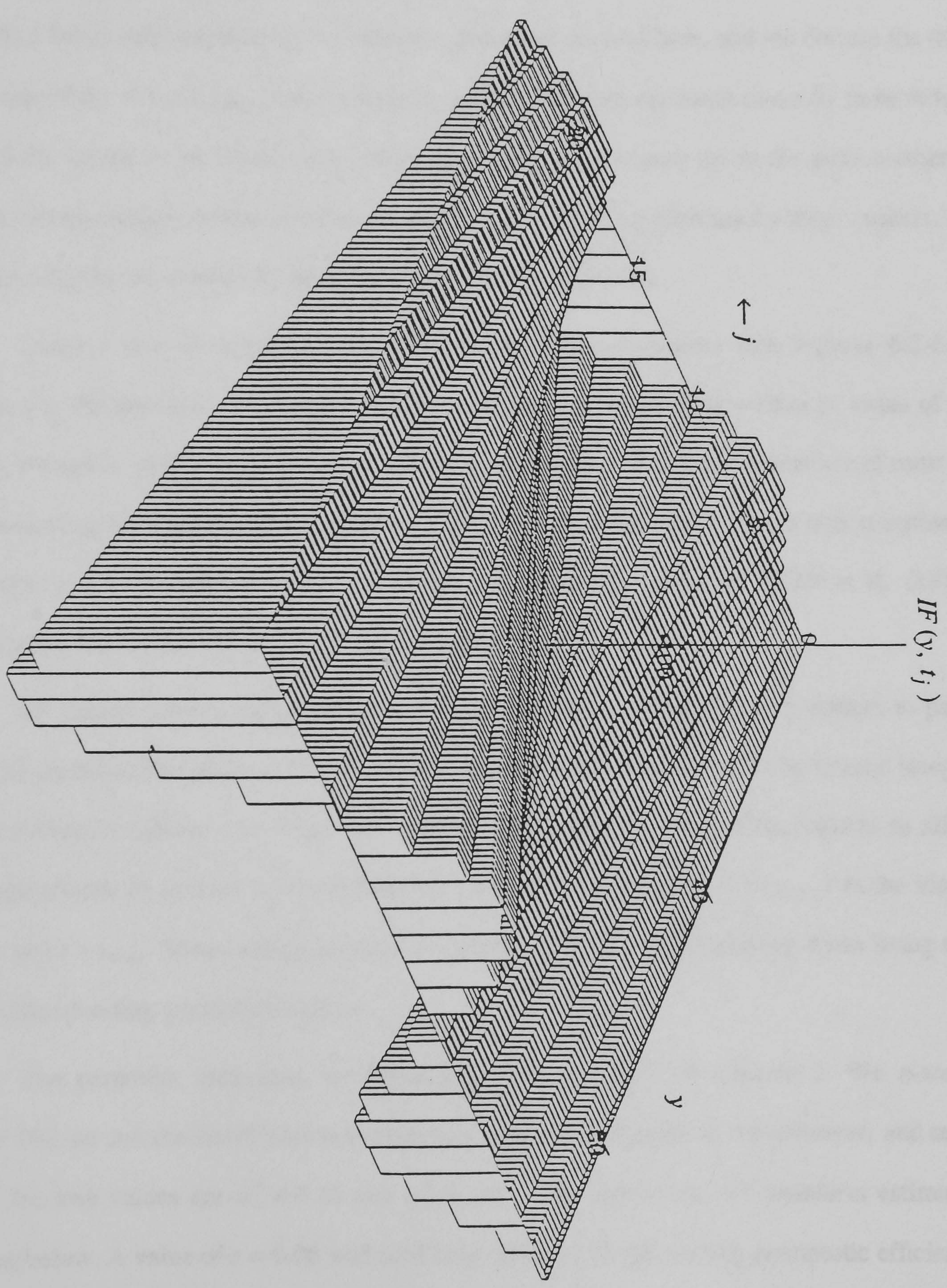


Figure 6.12. Influence Surface for Least-Squares Estimator of α_3.



developed an Ordered Simultaneous Estimation Procedure (OSEP), which requires that the data be partitioned into absorption, peak concentration and elimination phases. Three estimating equations are generated by selecting points from each of these regions, which may be solved for all possible triples. This produces sets of parameter estimates, with the OSEP estimates derived as the median of these. This is the fundamental reason for the robustness of their procedure, in that estimates are derived as medians, giving protection against outlying values. The conclusions they reached are closely matched by the influence functions derived here, and we discuss the detail for the case of $\hat{\alpha}_1$. For $t < t_{max}$ (see Table 6.2), outliers less than the mean cause $\hat{\alpha}_1$ to be reduced, as noted by Rodda et al. Their study only encompasses the region up to the peak concentration, where small outliers tend to increase $\hat{\alpha}_1$, with a corresponding decrease for large outliers. Similar discussions for the remaining estimators may also be conducted.

Turning now to a comparison with the transform estimators (see Figures 6.2-6.4), the influence surfaces are very similar. In particular, there is no obvious winner in terms of robustness, at least in global terms. However, the influence functions employed here are of most use for investigating local perturbations. That is, outliers in the vicinity of the mean concentration curve. In order to aid this interpretation, we conduct a study similar to that of Rodda et al. (1975), but we extend it to outliers beyond the peak concentration.

Of interest here is the response of the least-squares and transform estimators to particular outlier patterns. We generate "ideal" data as the mean values rounded to the nearest integer, and then introduce outliers into these data in three groups of two, to give six patterns in all. These groups consist of outliers above and below the mean in the regions $t < t_{max}$, t in the vicinity of t_{max} and $t > t_{max}$. These outlier patterns are tabulated below, with sampling times being those of the corresponding quadrature scheme.

The particular estimators we investigate here are those of Chapter 5. We assume that $\alpha_3 = 100$, so that the initial dose is known, and estimate the transition intensities α_1 and α_2, given that the true values are $\alpha_1^* = 0.10$ and $\alpha_2^* = 0.05$. The results for the transform estimators are given below. A value of $s = 0.06$ was used here, selected by optimizing asymptotic efficiency

Table 6.3. Outlier Patterns Used in Robustness Study.

t	"Ideal"	Outlier Pattern 1	2	3	4	5	6
0.0590	1	1	1	1	1	1	1
0.311	3	3	3	3	3	3	3
0.766	7	7	7	7	7	7	7
1.43	13	13	13	13	13	13	13
2.29	19	8 *	30 *	19	19	19	19
3.37	26	26	26	26	26	26	26
4.67	33	33	33	33	33	33	33
6.18	39	39	39	39	39	39	39
7.93	44	44	44	44	44	44	44
9.91	48	48	48	48	48	48	48
12.1	50	50	50	30 *	70 *	50	50
14.6	50	50	50	50	50	50	50
17.4	49	49	49	49	49	49	49
20.5	46	46	46	46	46	46	46
23.9	42	42	42	42	42	42	42
27.6	38	38	38	38	38	38	38
31.8	32	32	32	32	32	32	32
36.4	27	27	27	27	27	27	27
41.5	22	22	22	22	22	11 *	33 *
47.2	17	17	17	17	17	17	17
53.6	13	13	13	13	13	13	13
61.1	9	9	9	9	9	9	9
70.0	6	6	6	6	6	6	6
81.5	3	3	3	3	3	3	3

* - Aberrant Value.

versus least-squares (see Figures 5.28 & 5.34).

We consider the results of the robustness study in conjunction with the influence surfaces for the estimators, which follow in Figures 6.13 & 6.14 below. Note that the method of transform construction employed is that based on quadrature; earlier work tells us that conclusions for the Riemann-sum technique will be very similar.

Figure 6.13. Influence Surface for Transform Estimator of α_1.

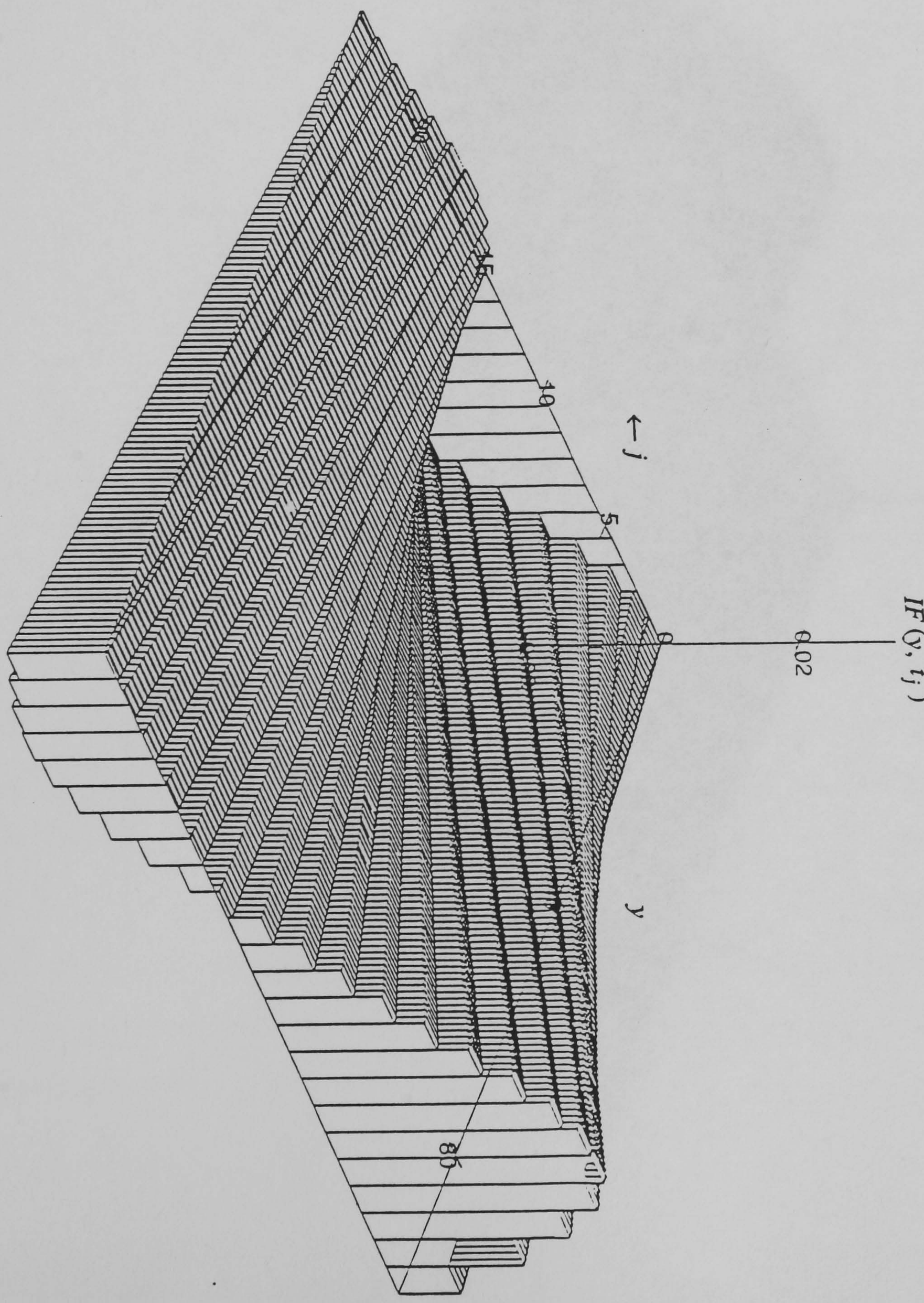
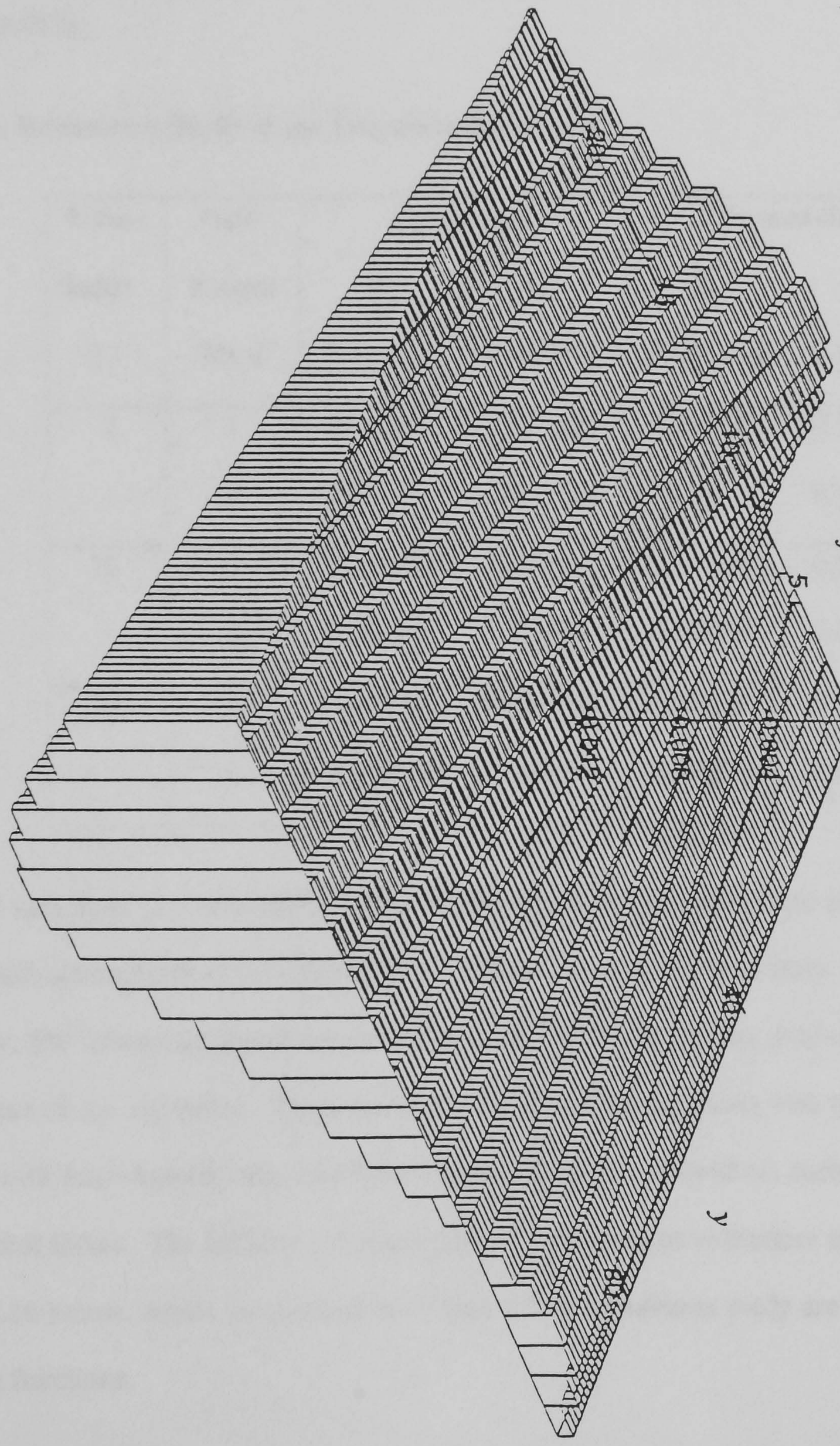


Figure 6.14. Influence Surface for Transform Estimator of α_2.



The results of the robustness study are summarized in Table 6.4 below. For $\hat{\alpha}_2$, we see that these results follow the influence surface very closely in each of the three regions. For $t < t_{max}$, small outliers can have a negative effect, whilst large outliers tend to increase the value of $\hat{\alpha}_2$. Around t_{max} small outliers have an obvious positive impact, with large outliers tending to reduce $\hat{\alpha}_2$. Beyond t_{max} the effect on this estimator ranges from little or none to negative for extreme outliers. The correspondence between the influence surface and the robustness study is also clear in the case of $\hat{\alpha}_1$.

Table 6.4. Robustness Study of the Transform Estimators.

Time-Index (j)	Data-Pattern	Quadrature $\hat{\alpha}_1$	Quadrature $\hat{\alpha}_2$	Riemann-Sum $\hat{\alpha}_1$	Riemann-Sum $\hat{\alpha}_2$
	"Ideal"	0.098597	0.050035	0.099145	0.049718
5	1	0.093811	0.049757	0.094368	0.049450
	2	0.10360	0.050268	0.10414	0.049943
11	3	0.092331	0.051593	0.092664	0.051332
	4	0.10516	0.048504	0.10595	0.048136
19	5	0.10075	0.051951	0.10119	0.051524
	6	0.096590	0.048191	0.097230	0.047977

We turn now to a robustness study of least-squares. The numerical minimization required was carried-out using NAG routine E04JAF, as in Chapter 5, with results summarized in Table 6.5 below. The numerical algorithm experienced difficulties in places, although IFAIL=3 is a very mild failure of the algorithm. The seriousness of the failure increases with the value of IFAIL. In contrast with least-squares, the transform-based estimators suffered no such problems since they take explicit forms. The influence surfaces for the least-squares estimators are depicted in Figures 6.15 & 6.16 below. Again we see that the results of the robustness study are predicted well by the influence functions.

Table 6.5. Robustness Study of the Least-Squares Estimators.

Time-Index	Data Pattern	$\hat{\alpha}_1$	$\hat{\alpha}_2$	IFAIL
	"Ideal"	0.10029	0.049976	3
5	1	0.096823	0.049880	6
	2	0.10407	0.050056	
11	3	0.092174	0.051393	
	4	0.10847	0.048517	
19	5	0.10270	0.051868	5
	6	0.097914	0.048121	

To conclude this chapter we turn to a comparison between the results for the transform estimators and least-squares. Given the influence surfaces, along with the additional information of the robustness studies, we are able to reach concrete conclusions. For $t < t_{max}$, transforms are more sensitive to outlying observations. However, beyond this region transforms are superior. In practical terms, the transform-based estimators should be regarded with some suspicion if an outlier is observed prior to the peak concentration. Beyond this point, though, these estimators seem to be reasonably insensitive to such observations. The reason for this behaviour is clear. Basing estimation on a transform of the conditional expectation $n_2(t; \underline{\alpha})$ introduces a non-robust functional, having an unbounded Influence Function. This component is down-weighted as t increases because we choose the Laplace transform in our work.

Figure 6.15. Influence Surface for Least-Squares Estimator of α_1.

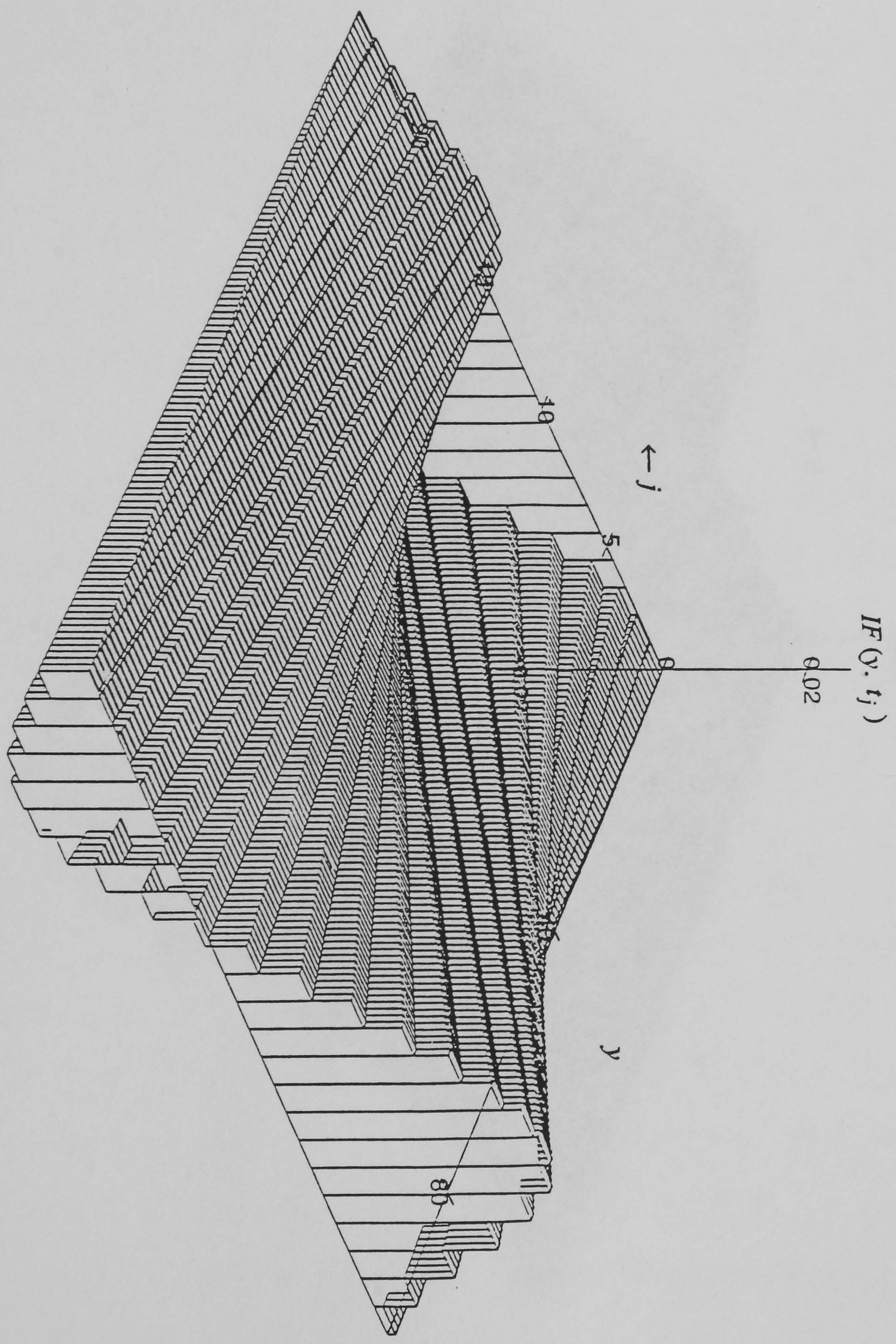
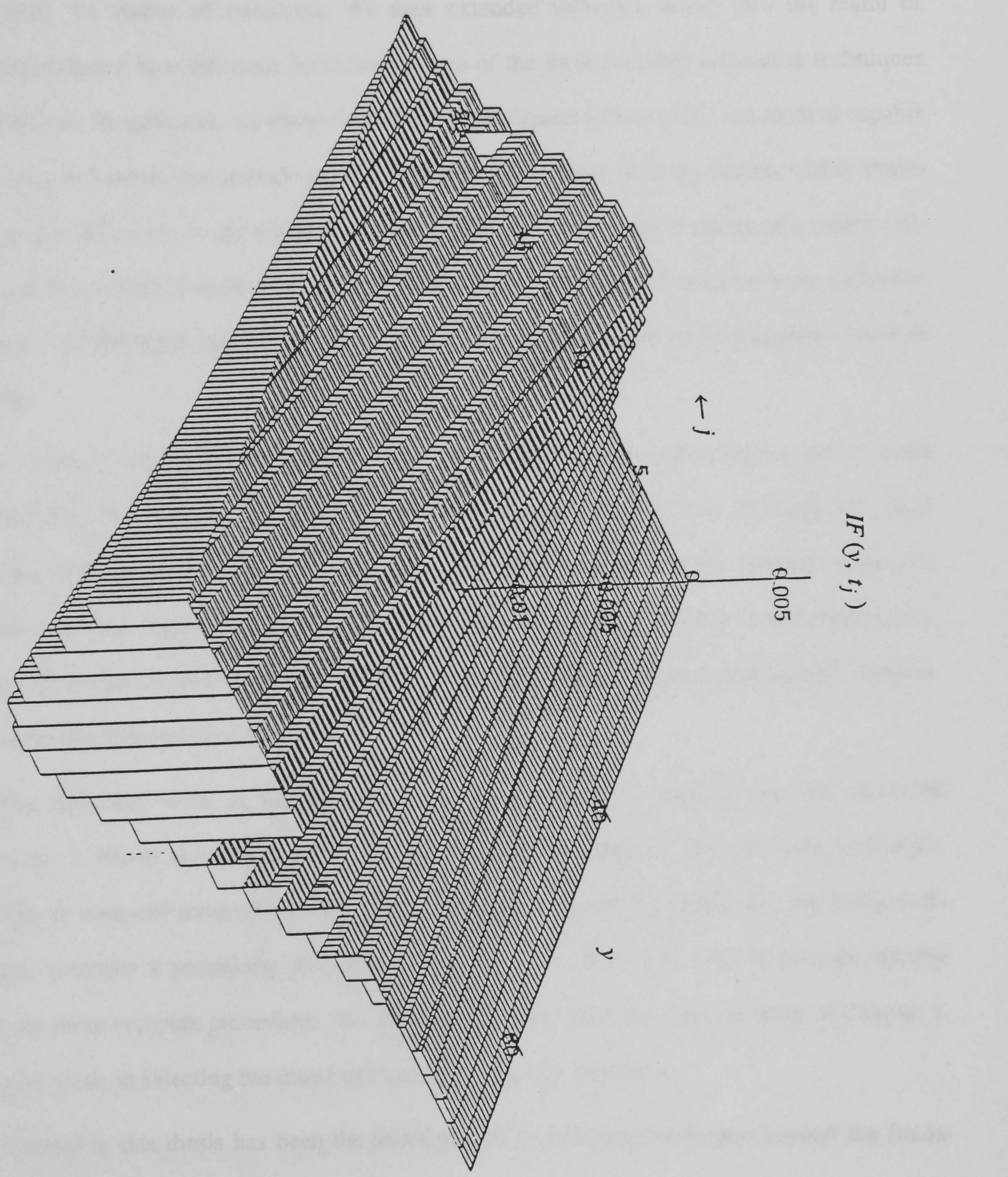


Figure 6.16. Influence Surface for Least-Squares Estimator of α_2.



CHAPTER 7:

CONCLUSIONS AND FUTURE WORK.

Given the structure of this thesis, it is natural that this concluding chapter should deal with the non-indexed and indexed cases essentially separately.

Beginning with the former, our first important result was that influence-bounded robustness follows from the choice of transform. We then extended influence theory into the realm of transforms, showing how influence functions for two of the most common estimation techniques may be derived. In particular, we showed that Integrated-Squared-Error (ISE) is a method capable of producing estimators that provide substantial protection against outlying values, whilst attaining acceptable efficiency in the absence of contamination. This is a key property of a robust estimator, and future work should seek to discover if this is true of ISE estimators more generally. The analysis of ISE conducted in this thesis strongly suggests that such an investigation would be profitable.

In contrast, the method-of-moments was shown in a number of examples to be rather inefficient. This is especially marked, as we saw in the Lagged-Normal work of Chapter 4, when estimation is base on modified transforms. Of particular interest are the moment equations obtained by taking successive derivatives of the transform employed. We observed a result analogous to that of conventional moment estimation, in that estimation becomes increasingly unstable as the order of differentiation increases.

The influence work in the context of moment estimation suggests that the observed inefficiency is due to an over-emphasis on obtaining robust estimators. ISE, however, retains the flexibility to trade-off some of this robustness for added efficiency. Despite this, the method-of-moments provides a potentially simple method that could at least be used to provide starting values for more complex procedures. To facilitate this it seems clear, from the work of Chapter 3, that more work on selecting the transform variable would be beneficial.

Central to this thesis has been the development of influence techniques beyond the funda-

mental definition of an IF, which has been shown to be a powerful and flexible tool. Given a sensible theoretical framework, it is also a quantity that is relatively easy to obtain, even when the corresponding statistic is not. An example of this was provided in Chapter 3, where we saw that it was possible to obtain influence functions for estimators defined by quite general estimating equations. The IF should find many useful applications in the future.

We turn now to the consideration of indexed random variables, and the more general area of stochastic modelling. We have employed the Laplace transform in this thesis, and fundamental to this work is the construction of an empirical Laplace transform (ELT). Two distinct approaches have been investigated, and their performance as competing methods has been broadly the same. However, their application in future practice may well be very different. When designing an experiment in advance we would prefer to employ quadrature, taking advantage of the ability to choose the sampling points. If such experimental design is not possible, such as when data have already been collected and are presented for analysis, then we would use the Riemann-sum technique.

We have undertaken quite a detailed study of these methods, largely in the context of the One-Compartment Open (OCO) model. Despite this, there are a number of important lines of investigation for future study. With regards quadrature, Sagar et al. (1992) show that its performance can vary considerably depending on the algebraic form of the integral to be approximated. They employed quadrature to estimate a particular integral, examining three different parameterizations of the integrand. It was observed that one of these yielded a distinctly superior approximation, and similar findings may also apply to the stochastic approximations of interest to us here. Turning to the Riemann-sum technique, we made no progress in applying end-corrections to this transform estimator in the context of the OCO model. These were possible for the Quantal Assay model, and yielded significant improvements over the uncorrected ELT. As such, it would likely be beneficial if some progress could be made towards applying end-corrections in general. The prospects do not, however, seem encouraging. A further point to note is that when comparing the quadrature and Riemann-sum techniques we have employed the quadrature-based sampling

points. It would be of interest to see if similar conclusions are reached when the Riemann-sum estimator is calculated using different sampling points.

In section 3.1.3 of Chapter 5 we proposed an alternative quadrature technique to the more usual, Gauss-Laguerre, rule. The standard approach is based on a polynomial approximation, whereas the proposed approximation is formed as a mixture of negative-exponential terms. This may well be more appropriate for models defined, as are compartment models, in terms of negative exponentials. This is a proposition worthy of further investigation, but does have a potential drawback. The sampling points and quadrature weights for this rule are not readily available, although these could be obtained fairly easily using a symbolic algebra package, such as MAPLE for example.

An important motivation for considering transform methods is the desire to obtain explicit parameter estimators. We have seen that this is possible for the OCO model, whether the initial dose is known or not. Besides explicit estimation, the transform-based estimators also suffered less from numerical problems than did ordinary least-squares. The work of Chapter 5 also showed that these estimators are capable of achieving very high efficiency compared to least-squares. However, the scope for extending these techniques beyond two compartments seems limited. Time does not permit further inquiry in this thesis, but it is clearly important that this question be investigated.

Turning to influence methods for the indexed case, we were able to develop techniques for gathering useful qualitative information based on a very simple framework. The results so obtained were, however, somewhat mixed. The reason for this is that estimation is based on a conditional expectation, a functional whose Influence Function is unbounded. The choice of the Laplace transform in our work tends to damp-down this component, especially for larger times. The sensitivity of the transform estimators to outlying observations can be reduced by choosing a larger value for the transform variable. However, our early work on transform construction showed that the ELT becomes progressively more biased as the transform variable becomes larger. As such, this does not represent a viable means of improving the robustness of the

transform estimators, and remains a topic for further study.

Future applications of transform methods in stochastic modelling should perhaps be focused where transforms occur naturally; for example, Queueing Theory and Insurance Mathematics. This latter suggestion is discussed by Teugels (1990), where the calculation of an empirical Laplace transform is required, although it is of the non-indexed variety. Another key application is to parameter estimation when a probability density is not easily found, as is the case when random variables are observed as convolutions. This is an area which we investigated in Chapter 4, but we are not limited to convolutions, as is clear from our discussion of Wise (1989) in the first chapter.

APPENDIX I : Parseval's Theorem and an Application to ISE.

Firstly, define the Fourier transform

$$\tilde{f}(s) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{isx} f(x) dx$$

of the function $f(x)$, along with the inner-product

$$< f, g > = \int_{-\infty}^{\infty} f(x) g(x)^* dx ,$$

where * denotes complex conjugate. The conclusion of Parseval's Theorem is that

$$< f, g > = < \tilde{f}, \tilde{g} > .$$

In terms of characteristic functions, when f and g are both probability density functions, we have, in obvious notation,

$$\int_{-\infty}^{\infty} f(x) g(x) dx = \int_{-\infty}^{\infty} (2\pi)^{-1} \phi_f(s) \phi_g(s)^* ds ,$$

since g is a real function. Further, if $f = g$ then

$$\int_{-\infty}^{\infty} |\phi_f(s)|^2 ds = 2\pi \int_{-\infty}^{\infty} [f(x)]^2 dx .$$

Given constants a and b, then the transform of $af + bg$ is $a\tilde{f} + b\tilde{g}$. We therefore obtain

$$\int_{-\infty}^{\infty} |\phi_f(s) - \phi_g(s)|^2 ds = 2\pi \int_{-\infty}^{\infty} [f(x) - g(x)]^2 dx .$$

APPENDIX II : Estimating Equations That Result From Allowing $\underline{s} \rightarrow s\underline{1}$.

The general case is most easily arrived at by consideration of the smallest relevant example, the two parameter case, where the estimating equations are

$$\hat{G}(s_i) = G(s_i; \hat{\Theta}) \quad i=1,2.$$

For the single s-value approach, we have to consider the behaviour of these equations as s_1 and s_2 approach some common value, s say. In the neighbourhood of s, the estimating equations may be rewritten as

$$\begin{aligned} \hat{G}(s) &= G(s; \hat{\Theta}) \\ \hat{G}(s+\delta) &= G(s+\delta; \hat{\Theta}), \end{aligned} \tag{A2.1}$$

for small $|\delta|$.

Now, G and $\hat{G}$ are both many times differentiable, so we may write (A2.1) as

$$\hat{G}(s)+ \delta\hat{G}'(s) = G(s; \hat{\Theta})+ \delta G'(s; \hat{\Theta})+ O(\delta^2).$$

We already have that $\hat{G}(s) = G(s; \hat{\Theta})$, so that (A2.1) reduces to

$$\hat{G}'(s) = G'(s; \hat{\Theta})+ O(\delta).$$

Allowing $\delta \rightarrow 0$, the estimating equations become

$$\begin{aligned} \hat{G}(s) &= G(s; \hat{\Theta}) \\ \hat{G}'(s) &= G'(s; \hat{\Theta}). \end{aligned}$$

In the general, p-parameter, case, we propose that the estimating equations become

$$\hat{G}^{(i-1)}(s) = G^{(i-1)}(s; \hat{\Theta}) \quad i=1,...,p,$$

where the superscripts denote the order of differentiation.

Proof - By Induction.

We have already demonstrated the result for $p = 2$, so we next assume it to be true for $p = k$.

Thus, the estimating equations are

$$\hat{G}^{(i-1)}(s) = G^{(i-1)}(s; \hat{\Theta}) \quad i=1,...,k.$$

Consider the case $p = k+1$. By the induction assumption, the estimating equations are

$$\begin{aligned} \hat{G}^{(i-1)}(s) &= G(s; \hat{\Theta}) \quad i=1,...,k, \\ \hat{G}(s_{k+1}) &= G(s_{k+1}; \hat{\Theta}). \end{aligned}$$

In the neighbourhood of s, we may write $s_{k+1} = s + \delta_{k+1}$, for small $|\delta_{k+1}|$. The $(k+1)$th estimating equation then becomes

$$\hat{G}(s+\delta_{k+1}) = G(s+\delta_{k+1}; \hat{\Theta}) .$$

Applying a Taylor-series expansion to both sides we obtain

$$\sum_{i=1}^{k} \hat{G}^{(i-1)}(s)\frac{(\delta_{k+1})^{i-1}}{(i-1)!} + \hat{G}^{(k)}(s)\frac{(\delta_{k+1})^{k}}{k!} = \sum_{i=1}^{k} G^{(i-1)}(s; \hat{\Theta})(\delta_{k+1})^{i-1} +$$
$$G^{(k)}(s; \hat{\Theta})\frac{(\delta_{k+1})^{k}}{k!} + O((\delta_{k+1})^{k+1}) .$$

Employing estimating equations 1 to k, the $(k+1)$th equation becomes

$$\hat{G}^{(k)}(s) = G^{(k)}(s; \hat{\Theta}) + O(\delta_{k+1}) .$$

The desired result follows on allowing $\delta_{k+1} \rightarrow 0$.

APPENDIX III : MAPLE Routines Used in Chapter 4.

The function of each of these programs is described in context during Chapter 4. We note here that quite sophisticated calculations may be undertaken which employ only a few lines of code.

A3.1

```
m := exp(0.5*s^2)/(1-s);                          Define the m.g.f.
m1 := diff(m,s);                                  First derivative of m.g.f.
ifpsi := array(1..4);                             Influence vector for transforms.
ifpsi[1] := exp(s*X)/ M *(X-m1/m);
   for i from 2 to 4 do
   ifpsi[i] :=diff(ifpsi[i-1],s):                 Fill-out rest of vector.
   od;
```

A3.2

```
eyl := array(1..5);                               Vector of expectations.
eyl[1] := m;
 for i from 2 to 5 do
   eyl[i] := diff(eyl[i-1],s);                    Build-up vector iteratively.
 od;
```

A3.3

```
read 'ifpsi_poly.sav'                             Read in the polynomial parts.
prod := array(symmetric,1..4,1..4);
   for i from 1 to 4 do                           Fill-in upper-triangle.
   for j from i to 4 do
      prod[i,j] := ifpsi[i]*ifpsi[j]:             Cross-product for [i,j]th element.
   od;
   od;
```

A3.4

```
eyl2 := array(1..9);
m2s := exp(2*s^2)/ (1-2*s);                       Define M(2s).
eyl2[1] := m2s;
   for i from 2 to 9 do
     eyl2[i] := 0.5*diff(eyl2[i-1],s):            Build-up eyl2 iteratively.
   od;
```

A3.5

```
read 'var_psi';                          Reads in var( Ψ̂ ).
with(linalg);                            Enable linear algebra facilities.
gv := array(1..30);                      Array to hold gv-values.
sv := array(1..30);                      Values of s.
   for i from 1 to 30 do
   s := -1.00 + (i-1)*0.05:              Scan interval in steps of 0.05.
   sv[i] := s:
   detvar := det(var_psi):               Calculate | var( Ψ̂ ) |
   gv[i] := (1-s)^16*detvar:
od;
```

References.

Abramowitz, M. & Stegun, I.A. (1970). Handbook of Mathematical Functions. Dover.

Alsihassan, W.A., Saidel, G.M. & Durand, D. (1992). Estimation of electrotonic parameters of neurons using an inverse Fourier transform technique. IEEE Transactions on Biomedical Engineering 39, 5, 493-501.

Allen, D.M. (1983). Parameter estimation for non-linear models with emphasis on compartmental models. Biometrics 39, 629-638.

Barndorff-Nielsen, O.A. & Cox, D.R. (1989). Asymptotic Techniques for use in Statistics. Chapman and Hall.

Barnett, V.D. & Lewis, T. (1978). Outliers in Statistical Data. Wiley.

Benichou, J. & Gail, M. (1990). Variance calculations and confidence intervals for estimates of the attributable risk based on logistic models. Biometrics **46**, 991-1003.

Bryant, J.L. & Paulson, A.S. (1983). Estimation of mixing proportions via distance between characteristic functions. Comm.Stat. A, **12**, 1009-29.

Brockwell, P.J. & Liu, J. (1992). Estimating the noise parameters from observations of a linear process with stable innovations. J.Stat.Plan.Inf. To appear.

Campbell, E.P. (1992). Influence for empirical transforms. Submitted for publication.

Chan, K.S. & Tong, H. (1990). Some comments on marginal distribution and autoregressive modelling. Technical Report.

Cook, R.D. & Weisberg, S. (1982). Residuals and Influence in Regression. Chapman and Hall: London.

Cox, D.R. & Hinkley, D.V. (1986). Theoretical Statistics. Chapman and Hall: London.

Critchley, F. (1985). Influence in principal components analysis. Biometrika 72, 3, 627-36.

Critchley, F. & Vitiello, C. (1990). On the influence of observations on misclassification probability estimates in linear discriminant analysis. University of Warwick Technical Report 204.

Crowder, M.J. (1986). On consistency and inconsistency of estimating equations. Econometric Theory 2, 305-330.

Crowder, M.J. (1987). On linear and quadratic estimating functions. Biometrika 74, 591-7.

Csörgö, S. & Heathcote, C.R. (1984). Testing by the empirical characteristic function. Technical Report No. 41, University of Ottawa.

Cummings, A. (1983). Parameter estimation using Laplace and Fourier transforms. Unpublished Diploma Project : University of Kent.

Davis, G.C. & Kutner, M.H. (1976). The lagged normal family of p.d.f.s applied to indicator-dilution curves. Biometrics 32, 669-675.

Delillo, N.J. (19882). Advanced Calculus with Applications. Collier-Macmillan : New York.

Epps, T.W. & Pulley, L.B. (1986). A test of exponentiality vs. monotone-hazard alternatives derived from the empirical characteristic function. J.R.Statist. B 48, 2, 206-213.

Faddy, M.J. (1992). A structured compartmental model for drug kinetics. To appear: Biometrics.

Fama, E.F. (1963). Mandelbrot and the stable paretian hypothesis. Journal of Business 36, 420-429.

Fama, E.F. (1965). Behaviour of stock market prices. Journal of Business 38, 34-105.

Feuerverger, A. & McDunnough, P. (1981). On the efficiency of empirical characteristic function procedures. J.Roy.Statist.Soc. B, 43, 20-27.

Feuerverger, A. & McDunnough, P. (1984). On statistical transform methods and their efficiency. Can.J.Statist. 12, 4, 303-317.

Feuerverger, A. & Mureika, R.A. (1977). The empirical characteristic function and its applications. Annals of Statistics 5, 1, 88-97.

Finney, D.J. (1971). Probit Analysis. 3rd edition. Cambridge University Press.

Frome, E.L. & Yakatan, G.J. (1980). Statistical estimation of the pharmacokinetic parameters in the one-compartment open model. Comm.Stat.Sim.Comp. B, 9, 201-222.

Hall, P. & Johnstone, I. (1992). Empirical functionals and efficient smoothing parameter selection. J.R.Statist.Soc. B, **54**, 2, 475-530 (including discussion).

Hampel, F.R, Ronchetti, E.M., Rousseeuw, P.J. & Stahel, W.A. (1986). Robust Statistics : The Approach Based on Influence Functions. Wiley.

Heathcote, C.R. (1977). The integrated squared error estimation of parameters. Biometrika **64**, 2, 255-64.

Heathcote, C.R. (1978). On parametric density estimators. Proceedings of Conference in Honour of Prof. J.E. Moyal. Adv.Appl.Prob. **10**, 735-740.

Huber, P.J. (1981). Robust Statistics. Wiley.

Katz, D. & D'Argenio, D.Z. (1983). Experimental design for estimating integrals by numerical quadrature, with applications to pharmacokinetic studies. Biometrics 39, 3, 621-628.

Kemp, A.W. & Kemp, C.D. (1987). A rapid and efficient estimation procedure for the negative binomial distribution. Biom.J. **29**, 865-873.

Kodell, R.L. & Matis, J.H. (1976). Estimating the rate constants in a two-compartment stochastic model. Biometrics 32, 377-400.

Koutrouvelis, I.A. (1980). Regression-type estimation of the parameters of stable laws. J.Am.Statist.Assoc. 75, 918-28.

Koutrouvelis, I.A. (1981). An iterative procedure for the estimation of the parameters of stable laws. Comm.Staist. B **10**, 17-28.

Koutrouvelis, I.A. (1982). Estimation of location and scale in Cauchy distributions using the empirical characteristic function. Biometrika **69**, 1, 205-13.

Laurence, A.F. & Morgan, B.J.T. (1987). Selection of the transformation variable in the Laplace transform method of estimation. Austral.J.Statist. **29**, 2, 113-127.

Laurence, A.F., Morgan, B.J.T. & Tweedie, R.L. (1986). Parameter-estimation in non-linear models using Laplace transforms. Unpublished manuscript.

Leedow, M.I. & Tweedie, R.L. (1983). Weighted area techniques for the estimation of the parameters of a growth curve. Aust.J.Statist. 2, 310-320.

Macheras, P., Reppas, C. & Symillides, M. (1992). Fraction of the bioavailable dose remaining in the body at the time of peak plasma concentration in a linear, open, one-compartment model. J.Pharma.Sci. **81**, 1, 110-112.

Mandelbrot, B. (1963). The variation of speculative prices. Journal of Business **36**, 394-419.

Maple User Guide (1989). Courtesy of the Liverpool University Algebra Support Group.

Matis, J.H. & Hartley,H.O. (1971). Stochastic compartmental analysis : model and least-squares estimation from time series data. Biometrics **27**, 77-102.

Matis, J.H., Wehrly, T.E. & Ellis, W.C. (1989). Some generalized stochastic models for digesta flow. Biometrics **45**, 3, 703-720.

McCullagh, P. & Nelder, J.A. (1989). Generalized Linear Models. Chapman and Hall.

Miller, R.G. & Halpern, J. (1980). Robust estimators for quantal bioassay. Biometrika **67**, 103-110.

Morgan, B.J.T. (1982). Parameter estimation via Laplace transform estimation. Proceedings of COMPSTAT'82, Part II, pp 195-96.

Pack, P. & Jolliffe, I.T. (1992). Influence in correspondence analysis. Appl.Statist. **41**, 2, 365-380.

Paulson, A.S., Holcomb, E.W. & Leitch, R.A. (1975). The estimation of the parameters of the stable laws. Biometrika **62**, 163-170.

Paulson, A.S. & Nicklin, E.H. (1983). Integrated distance estimators for linear models applied to some published data sets. Appl.Stats. **32**, 1, 32-50.

Phillips, C. & Cornelius, B. (1986). Computational Numerical Methods. Ellis Horwood.

Quandt, R.E. & Ramsey, J.B. (1978). Estimating mixtures of normal distributions and switching regressions. J.Am.Statist.Assoc. **73**, 730-752.

Rabinowitz, P. & Davis, P.J. (1984). Methods of Numerical Integration. Academic Press.

Rodda, B.E., Sampson, C.B. & Smith, D.W. (1975). The One-compartment Open Model: Some statistical aspects of parameter estimation. Appl.Statist. **24**, 3, 309-318.

Sagar, R.P., Schmider, H. & Smith Jr., V.H. (1992). Evaluation of Fourier transforms by Gauss-Laguerre quadratures. J.Phys.A:Math.Gen. **25**, 189-195.

Schmidt, P. (1982). An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions. Econometrica **50**, 501-524.

Schuh, H.J. & Tweedie, R.L. (1979). Parameter estimation using transform estimation in time-evolving models. Math.Biosciences **45**, 37-67.

Seber, G.A.F. & Wild, C.J. (1989). Nonlinear Regression. Wiley.

Shah, B.K. (1976). Data analysis problems in the area of pharmacokinetics research. Biometrics 32, 145-157.

Teugels, J.L. (1990). Topics in insurance mathematics. Invited paper presented at the "Conferência em Estatistica e Optimizacão", Troia, Portugal Dec. 3-5, 1990.

Titterington, D.M., Smith, A.F.M. & Makov, U.E. (1985). Statistical analysis of finite mixture distributions. Wiley.

Urfer, W. (1992). Estimation of toxicokinetic parameters using a two-compartment model. Information Bulletin 2, Portugese Society of Statistics.

Wagner, J.G. (1967). Use of computers in pharmacokinetics. Clin. Pharmacology and Therapeutics 8, 201.

Wald, J,A., Salazar, D.E., Cheng, H. & Jusko, W.A. (1991). Two-compartment basophil trafficking model for methylprednisolone pharmacodynamics. Journal of Pharmacokinetics and Biopharmaceutics 19, 5, 521-536.

Wise, K.W. (1989). Statistical Aspects of Estimating Environmental Plutonium. Unpublished PhD Thesis. La Trobe University.