ON THE ADEQUACY OF THE

SARAGAN APPROXIMATION TO THE NORMAL

IN ECONOMETRIC MODELS

By

Mohammad-Ali Kafaei

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1984

ABSTRACT

ON THE ADEQUACY OF THE
SARAGAN APPROXIMATION TO THE NORMAL
IN ECONOMETRIC MODELS

By

Mohammad-Ali Kafaei

The normal distribution is very commonly assumed in
econometric models, but it causes some problem in models for which the
likelihood function involves the normal c.d.f. Goldfeld and Quandt
introduced a class of distributions, called Sargan distributions, to be
used as an approximation to the normal, which have certain desirable
properties, notably an integrable c.d.f. We investigate the adequacy
of these Sargan distributions in approximating the normal. This is
done by measuring the "cost" resulting from the usage of the Sargan
distribution instead of the normal as the assumed distribution of the
error terms. The cost is defined in terms of the asymptotic bias (or
inconsistency) of the parameter estimates, when the errors are actually
normal. We prove that this "cost" is model dependent. There is no
cost in the linear regression model, but in other models (such as
sample selection models) there is a cost (a false distributional
assumption causes inconsistency), and the cost goes up with the degree
of censoring or truncation. We consider three different models for
both the realistic case of an unknown variance and the realistic case
of a known variance, considering both first and second order univariate

Sar

exa

to

biv

use

reg

tha

not

lar

ben

com

esp

mul

inv

Sargan distributions. We also define a bivariate Sargan distribution, examine its properties and investigate its adequacy as an approximation to the bivariate normal distribution. We provide tables for the bivariate Sargan and normal densities, compare them, and discuss the use of this Sargan distribution in a simple seemingly unrelated regression model.

The overall conclusion of this study is that if one believes that the true distribution of the error terms is normal, it really does not pay that much to, use the Sargan distribution. Especially in largely censored (or truncated) samples, the cost is more than the benefits (in terms of computational savings) gained. But the potential computational saving is generally higher in the bivariate (or especially the multivariate) case than in the univariate case, so multivariate Sargan distributions may be worthy of further investigation.

To the Memory of my father,

and to my mother, brother and sisters.

ACKNOWLEDGMENTS

I would like to express my sincere gratitute and appreciation to Professor Peter J. Schmidt, chairman of my dissertation committee for his invaluable advice, suggestions, comments and criticism as well as his encouragement and supports without which this study would have not been possible. I am also grateful to the other members of the committee, Professor Robert H. Rasche, professor Daniel S. Hamermesh, and Professor Byron W. Brown for their critical reading and useful Comments and suggestions.

I also would like to thank Mrs. Terie Snyder for typing the equations of the dissertation.

Most of all I will be indebted to the members of my family for their moral and financial support.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Introduction

1.1 Historical Background

The assumption of normality is very common in econometric and statistical work. For example, errors in regression models are often assumed to be normally distributed; hypotheses about means of random variables are tested under the assumption of normality: and in more complicated models, such as disequilibrium models or bivariate probit models, a multivariate normal distribution is assumed. To some extent, the assumption of normality can be justified by reliance on the central limit theorem. However, such justifications are seldom very rigorous, and it is probably fair to say that the frequent use of the normal distribution is in fact due to two other reasons. First, in many common models the assumption of normality is very convenient. For example, in the linear regression model, the maximum likelihood estimator under normality is least squares, which is simple to calculate. Also given normality exact finite sample tests of linear hypotheses are possible. Alternative error distributions would lead to more complicated estimation and testing procedures, which would generally be justified only asymptotically. Second, in many common models inferences based on normality are asymptotically robust to non-normality. Again using the linear regression model as an example,

1

tests which are exact in finite samples given normality are correct asymptotically given any non-normal distribution with finite mean and variance.

However, in other models the assumption of normality may not be very convenient. The cumulative distribution function of the normal distribution can not be expressed in closed form, and therefore the normality assumption may not be the most convenient in models for which the likelihood function contains the c.d.f. of the error distribution. This is so in a wide class of models involving censoring, truncation or selection of the dependent variable (e.g., the Tobit model). This is especially true in multi-equation models (e.g., a multi-market disequilibrium model), since the c.d.f. of the multivariate normal distribution can be very expensive to evaluate.

Such computational considerations led Goldfeld and Quandt(1981) to introduce a class of distributions, which they call "Sargan distributions", to be used as substitutes for the normal distribution in models for which estimation requires evaluation of the c.d.f. of the error distribution. The Sargan c.d.f. can be evaluated analytically, and its density is reasonably close to the normal density. Thus Sargan distributions may be reasonable candidates to use as an approximation to the normal distribution.

An obvious question to ask is how good an approximation to the normal distribution a Sargan distribution provides. One way to answer this question is simply to compare the densities (or c.d.f.'s) of the two distributions. Goldfeld and Quandt (1981, p. 145) provide in their Table 1 a comparison of the densities of $N(0,1)$ and of a

second-order Sargan distribution with mean zero, variance one, and the same density at zero as $N(0,1)$. The agreement is reasonable close, except in the tails. Missiakoulis (1983, p. 227) provides in his Table 1 the same results, plus the density of the first-order Sargan distribution with variance equal to one ($a=2$). As might be expected, the first-order Sargan distribution does not provide as good an approximation as does the second-order Sargan distribution.[1]

While such direct comparisons are informative and easy to make, they do not provide good evidence on the relevant statistical question, namely the effect on one's estimates or inferences of the use of such an approximation to the (hypothesized) true error distribution. This is a much harder question, in part because its answer clearly depends on the nature of the model in which the errors appear. For example, an incorrect assumption of normality does not cause bias or inconsistency of the parameter estimates in the linear regression model, but it does cause inconsistency in limited dependent variable models, such as the Tobit model. Conversely, the incorrect assumption that the errors have a Sargan distribution does not cause inconsistency in the linear regression model (as we prove later), but it does cause inconsistency in more complicated models such as the Tobit model. Indeed, it is an unfortunate fact that, for the class of models for which the Sargan assumption is convenient (models for which the likelihood function involves the error c.d.f.), the correctness of the assumed error distribution is vital, even asymptotically.

Some work has been done on the robustness of the normal maximum likelihood estimates to non-normality, in the Tobit model and some

related models. Goldberger (1980) and Arabmazar and Schmidt (1982) consider the Tobit model, the truncated version of the Tobit model, and the probit model, and consider the asymptotic bias (inconsistency) which results when the normality assumption is incorrect. The alternative distributions considered were Student's t, Laplace, and logistic, and the models considered contained only a constant term. The bias is sometimes substantial, so that this problem merits further attention. Missiakoulis (1983) has provided a similar analysis, for the probit model only, of the asymptotic bias resulting from an (incorrect) assumption that the errors have a Sargan distribution, when the normal is in fact correct.

In this thesis we consider some implications of the use of the Sargan distribution in econometric models. The plan of the thesis is given in section 1.3. Section 1.2 first defines Sargan distributions, and gives some of their properties, for later reference.

1.2 Univariate Sargan Distribution; Definitions and Properties

Goldfeld and Quandt (1981) define the family of Sargan densities as (equation 2.1);

$$(1.1) \quad f(x) = Ke^{-\alpha|x|} (1 + \sum_{j=1}^{P} \gamma_j \alpha^j |x|^j)$$

where $\alpha > 0$, $\gamma_j \geq 0$, $j=1,2,\ldots,P$ and P is the order of the density. Alternatively, in the notation of equation (1) in Missiakoulis (1983)

$$(1.2) \quad f(x) = \frac{D\alpha}{2} e^{-\alpha|x|} \sum_{j=0}^{P} \gamma_j \alpha^j |x|^j$$

with

$$(1.3) \quad D = (\sum_{i=0}^{P} \gamma_i \, i!)^{-1} = \frac{2K}{\alpha}$$

and $\gamma_0 = 1$.

Note that, as Goldfeld and Quandt and Missiakoulis mention, by setting $\gamma_j = 0$ for all j, $f(x)$ becomes a generalization of the Laplace density. Also, Goldfeld and Quandt (1981, p. 144) note that for $f'(x)$ to be continuous, one must impose $\gamma_1 = 1$.

The Sargan density is a symmetric, unimodal density function with X defined on $(-\infty , \infty)$. Its moment generating function can be derived as follows.

$$(1.4) \quad M(\theta) = E(e^{\theta x}) = \int_{-\infty}^{\infty} e^{\theta x} f(x)dx$$

Define:

$$A = \frac{D\alpha}{2} \sum_{i=0}^{P} \gamma_i \int_{-\infty}^{0} \alpha^i e^{(\alpha+\theta)x}(-x)^i dx$$

Using the fact (see Gronner and Hofreiter (1958), p. 55) that

$$(1.5) \quad \int_{0}^{\infty} x^n e^{-\alpha x}dx = \frac{n!}{\alpha^{n+1}} \qquad n > 0$$

$$A = \frac{D}{2} \sum_{i=0}^{P} \frac{\alpha^{i+1}\gamma_i \, i!}{(\alpha+\theta)^{i+1}}$$

and similary if we define

$$B = \frac{D\alpha}{2} \sum_{i=0}^{P} \gamma_i \int_{0}^{\infty} \alpha^i e^{-(\alpha-\theta)x} x^i dx = \frac{D}{2} \sum_{i=0}^{P} \frac{\alpha^{i+1}\gamma_i i!}{(\alpha-\theta)^{i+1}}$$

Thus

$$(1.6) \quad M(\theta) = A + B = \frac{D}{2} \sum_{i=0}^{P} \alpha^{i+1} \gamma_i \, i! \, [\frac{1}{(\alpha+\theta)^{i+1}} + \frac{1}{(\alpha-\theta)^{i+1}}]$$

Therefore, the rth moment of the Sargan distribution is

$$\mu_r(\theta=0) = \frac{D}{2} \sum_{i=0}^{P} \alpha^{i+1} \gamma_i (i!) \, [\frac{(-1)^r (i+1)(i+2)\ldots(i+r)}{\alpha^{i+r+1}}$$

$$+ \frac{(i+1)(i+2)\ldots(i+r)}{\alpha^{i+r+1}}]$$

$$= \frac{D}{2\alpha^r} \sum_{i=0}^{P} (i+r)! \gamma_i [(-1)^r + 1]$$

$$(1.7) \quad \mu_r = \begin{cases} 0 & \text{if } r \text{ is odd} \\\\ \frac{D}{\alpha^r} \sum_{i=0}^{P} (i+r)! \, \gamma_i & \text{if } r \text{ is even} \end{cases}$$

Equation (1.7) is different from equation (6) in Missiakoulis (1983) and from equation (2.3) of Goldfeld and Quandt (1981) by the factor γ_i^2.

The Sargan c.d.f. can be evaluated analytically as follows:

i) $x < 0$

$$F(x) = \int_{-\infty}^{x} \frac{D\alpha}{2} e^{\alpha y} \sum_{i=0}^{P} \gamma_i \alpha^i |y|^i dy$$

$$= \frac{D}{2} e^{\alpha x} \sum_{i=0}^{P} i! \gamma_i \sum_{j=0}^{i} \frac{(-\alpha x)^j}{j!}$$

ii) $x > 0$

$$F(x) = \int_{-\infty}^{0} \frac{D\alpha}{2} e^{\alpha y} \sum_{i=0}^{P} \gamma_i \alpha^i |y|^i dy + \int_{0}^{x} \frac{D\alpha}{2} e^{-\alpha y} \sum_{i=0}^{P} \gamma_i \alpha^i y^i dy$$

$$= \frac{D}{2} \sum_{i=0}^{P} \gamma_i i! - \frac{D}{2} \sum_{i=0}^{P} \gamma_i \alpha^{i+1} \sum_{j=0}^{i} \left[\frac{i! x^j e^{-\alpha x}}{j! \alpha^{i-j+1}} - \frac{i!}{\alpha^{i+1}} \right]$$

$$= D \sum_{i=0}^{P} \gamma_i i! - \frac{D}{2} e^{-\alpha x} \sum_{i=0}^{P} \gamma_i i! \sum_{j=0}^{i} \frac{(\alpha x)^j}{j!}$$

That is,

$$(1.9) \qquad F(x) = \begin{cases} \frac{D}{2} e^{\alpha x} \sum_{i=0}^{P} \gamma_i i! \sum_{j=0}^{i} \frac{(-\alpha x)^j}{j!} & \text{if } x < 0 \\\\ 1 - \frac{D}{2} e^{-\alpha x} \sum_{i=0}^{P} \gamma_i i! \sum_{j=0}^{i} \frac{(\alpha x)^j}{j!} & \text{if } x > 0 \end{cases}$$

The special cases of (1.9) for the first-order and second-order Sargan densities are defined in chapter 3 (equation 3.18) and chapter 4 (equation 4.13). The first-order and second-order Sargan densities and their moments are as follows:

Firs

(1.10

(1.1

Seco

(1.

(no

(1.

(1.

and

First-Order Case

(1.10) $\quad f(x) = \frac{\alpha}{4} e^{-\alpha|x|} (1+\alpha|x|)$

$\mu = 0$

$var(x) = \alpha^2 / 4$

(1.12) $\quad M(\theta) = \frac{1}{4} \left(\frac{\alpha}{\alpha+\theta} + \frac{\alpha}{\alpha-\theta} + \frac{\alpha^2}{(\alpha+\theta)^2} + \frac{\alpha^2}{(\alpha-\theta)^2} \right)$

Second-Order Case

(1.13) $\quad f(x) = \frac{\alpha e^{-\alpha|x|}}{4(1+\gamma_2)} (1+\alpha|x| + \gamma_2 \alpha^2 x^2)$

(note that $\gamma_1 = 1$ is already set)

$\mu = 0$

(1.14) $\quad var(x) = \dfrac{4(1+3\gamma_2)}{\alpha^2(1+\gamma_2)}$

(1.15) $\quad M(\theta) = \dfrac{\dfrac{\alpha}{\alpha+\theta} + \dfrac{\alpha}{\alpha-\theta} + \dfrac{\alpha^2}{(\alpha+\theta)^2} + \dfrac{\alpha^2}{(\alpha-\theta)^2} + \dfrac{2\alpha^3\gamma_2}{(\alpha+\theta)^3} + \dfrac{2\alpha^3\gamma_2}{(\alpha-\theta)^3}}{4(1+\gamma_2)}$

Equation (1.15) is different from equation (2.3) of Goldfeld and Quandt (1981) as mentioned before.

the

mod

asy

inc

mea

one

ver

Con

whi

all

(T

be

"1

Y_i

ob

1.3 Outline of the Dissertation

The purpose of this study is to investigate the adequacy of the Sargan distribution as an approximation to the normal in econometric models. The measure of adequacy is a simple one; namely, the asymptotic bias (inconsistency) which results from the use of an incorrect assumption about the error distribution. Obviously other measures of adequacy of the approximation would be possible, but this one is informative and has been used profitably elsewhere.

The models in which we will investigate this question are all versions of what Heckman (1976) calls sample selection models. Consider a hypothetical linear relationship

$$y_i^* = \beta' X_i + \epsilon_i \qquad i=1,2,\ldots,n$$

which would satisfy the usual ideal conditions if y_i^* were observed for all i. In the <u>censored regression model</u>, we observe

$$(1.16) \quad y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ \\ 0 & \text{if } y_i^* < 0 \end{cases} = \max(0, y_i^*) \quad i=1,\ldots,n$$

(This is the Tobit model if ϵ is normal.) Observations with $y_i \geq 0$ will be called "complete" observations while those with $y_i = 0$ will be called "limit" observations. In the <u>truncated regression model</u> we observe $y_i = y_i^*$ if and only if $y_i^* \geq 0$. In other words, we observe the complete observations but nothing about the limit observations (not even their

numbe

(1

sc we

model

assum

exten

we a

in th

distr

same

chap

secon

dist

doin

marg

some

inve

to t

model

number). Finally, in the <u>binary regression mobel</u> we observe

$$(1.17) \quad y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* < 0 \end{cases} \quad i=1,\ldots,n$$

so we essentially observe only the sign of y_i^*. (This is the probit model if ϵ is normal.)

In chapter 2 we investigate the robustness of the MLE's which assume normality to Sargan errors, in these three models. This is an extension of Goldberger (1981) and Arabmazar and Schmidt (1982), which we address mostly for the sake of completeness. Our main interest is in the opposite question, the robustness of MLE's which assume a Sargan distribution to normal errors. We investigate this question in the same three models plus the linear regression model. This is done in chapter 3 for first-order Sargan distributions and in chapter 4 for second-order Sargan distributions.

In chapter 5 we consider how to generalize the Sargan distribution to the multivariate case. We consider alternative ways of doing so, and define a particular bivariate distribution with Sargan marginal distributions as a bivariate Sargan distribution. We make some comparisons of its density to that of the bivariate normal, and investigate the robustness of the MLE's based on the bivariate Sargan to the bivariate normal in a simple seemingly unrelated regressions model.

Finally, chapter 6 contains our conclusions.

well

beca

were

cont

be c

only

nest

p⁻⁻-

the

dc r

Sarg

whic

tha

Footnotes

1) Missiakoulis also displays the densities of some higher-order Sargan distributions. These approximate the normal less well than the second-order Sargan distribution. This is possible because of the way the parameters of the higher-order distributions were chosen. In particular, whereas the Sargan density of order P contains parameters (a , $\gamma_2, \ldots, \gamma_p$), Missiakoulis assumes the γ 's to be chosen a priori (see his equation (9)) so that for any P there is only one unknown parameter (a). Thus his second-order case is not nested in his third-order case, and so forth. If we allowed the Pth-order case to have P free parameters, and chose these to maximize the quality of the approximation (however defined), obviously we could do no worse by increasing the order of the approximating distribution.

2) Using (1.7) to calculate a when $\gamma_2 = 1/3$, P= 2 (second-order Sargan), and variance = 1,

$$\mu_2 = \text{variance} = D(2 + 6\,\gamma_1 + 24\,\gamma_2)/\,a^2 = 1$$

$$a = \sqrt{6} = 2.44949 , \qquad (D=3/8)$$

which is different from that of Missiakoulis (1983, p. 227). (Note that $\gamma_1 = 1$ is required to ensure the uniqueness of the results.)

di

al

or

di

fi

me

n

t

i

a

(

Chapter 2

Robustness of Normal MLE's to Sargan Errors

2.1 Introduction

The normal distribution has been assumed to be the true distribution of the error terms in a variety of different models, although in many cases there has not been a very good reason to believe or substantiate this assumption. This popularity of the normal distribution stems partly from computational considerations and partly from the robustness of the estimates based on normality. In certain models, such as the linear regression model, the MLE's based on normality are unbiased, BLUE and consistent even if the distribution of the errors is non-normal; only for testing hypothesis in small samples is the normality assumption crucial. The casual use of the normality assumption in different models requires more attention. In this chapter we look at this issue in different models with a specific and true non-normal distribution of the error terms, namely first-order Sargan, as defined in chapter 1.

This attempt can be seen as an extension of Goldberger (1981), and Arabmazar and Schmidt (1982) , who raised the question of the robustness of the normal MLE's to non-normality in the censored (Tobit) and truncated regression models. Arabmazar and Schmidt concluded that "the bias from non-normality can be substantial", especially if the

13

varianc

few di

interes

the act

measure

(incons

likeli

Theref

normal

regres

Sargan

from

effici

trunc

secti

resul

param

chapt

Gold

Miss

cons

variance of the errors is not known. However, they considered only a few distributions (Laplace, logistic and Student's t), so it is interesting to perform a similar analysis under the new assumption that the actual distribution of the errors is Sargan. As in their work, the measure of robustness to be used here is the asymptotic bias (inconsistency) of the estimates.

In the case of linear regression model, the normal maximum likelihood estimate (which is OLS) is unbiased and consistent. Therefore in terms of our analysis, there is no cost in assuming normality in the linear regression case even if in fact that is not so.

Goldfeld and Quandt (1981) consider the case of linear regression with the residuals actually being distributed as first-order Sargan, but incorrectly assumed to be normal. One important result from their paper is the fact that efficiency is lost; the relative efficiency of OLS to the Sargan MLE is .84.

In this chapter, I will consider in turn the censored, truncated and binary (probit) dependent variable models. In each section, after defining the model and its estimators, I discuss the results of numerical calculations of the inconsistency of the estimated parameters. (These are tabulated at the end of this chapter.) The chapter ends with some concluding remarks.

In order to make the argument tractable and also comparable to Goldberger (1980), Arabmazar and Schmidt (1982), and Missiakoulis (1983), we assume that there is only one regressor, a constant term, in all the models under consideration here.

where

only

mean

is th

we as

distr

estim

we ca

the

need

give

(2.

2.2 The Censored Case (Tobit Model):

The model to be considered here is;

$$(2.1) \quad y_i^* = \mu + u_i , \quad i = 1, \ldots, n ,$$

where u_i are i.i.d with zero mean and variance σ^2. By assumption, we only observe $y_i = \text{Max}(0, y_i^*)$, not y_i^*. We wish to estimate μ and σ^2, the mean and variance of y^*. The only thing that remains to be specified is the true distribution of u_i. We ask the question of what happens if we assume the errors to be normally distributed when in fact they are distributed as first order Sargan. In other words, is the Tobit estimator robust to non-normality? In order to answer this question, we can use equation (10) of Arabmazar and Schmidt (1982), and calculate the asymptotic bias of the Tobit estimates of μ and σ^2. For this we need the first two truncated moments of the Sargan distribution (as given for the other distributions in their Table I); these are:

$$(2.2) \quad E(u|u > -x) = \begin{cases} \dfrac{e^{-\alpha x}(\alpha x^2 + 3x + 3/\alpha)}{4 - e^{-\alpha x}(2 + \alpha x)} & x > 0 \\[4mm] \dfrac{\alpha x^2 - 3x + 3/\alpha}{2 - \alpha x} & x < 0 \end{cases}$$

and,

Table

The t

The r

binar

this

to Ta

assur

the p

esti

the

The p

degre

the a

and,

$$(2.3) \quad E(u^2|u > -x) = \begin{cases} \dfrac{\dfrac{16}{\alpha^2} - 4e^{-\alpha x}[\dfrac{\alpha}{4}x^3 + x^2 + \dfrac{2}{\alpha}x + \dfrac{2}{\alpha^2}]}{4-e^{-\alpha x}(2+\alpha x)} & x > 0 \\\\ \dfrac{-\alpha x^3 + 4x^2 - \dfrac{8}{\alpha}x + \dfrac{8}{\alpha^2}}{2-\alpha x} & x < 0 \end{cases}$$

The estimated biases of the Tobit estimates are given in Table 2.1, for a variety of values of μ, under the heading "censored". The true value of α is assumed to be two, which corresponds to $\sigma^2 = 1$. The results for the truncated version of Tobit model and also for the binary case (probit), which we will discuss in the next sections of this chapter, are also tabulated in Table 2.1. This table is similar to Tables II-V of Arabmazar and Schmidt (1982).

We consider both the unrealistic case in which variance is assumed to be known (so that the estimated α is set equal to two), and the realistic case in which σ^2 is unknown, so that μ and α must both be estimated. We will consider first the case of known variance.

The results and the conclusions drawn from them are basically the same as those of Arabmazar and Schmidt for the t_{10} distribution. The bias is sometimes substantial, but it is heavily dependent upon the degree of censoring. That is, as the sample becomes largely complete, the asymptotic bias goes quickly to zero. This agrees with the fact

(as i

regre

compl

that

subst

depen

which

(more

as th

in th

That

terms

unkno

censo

bette

thoug

with

the v

less

the

censo

virtu

(as it must) that there is no asymptotic bias in the case of the linear regression model. For example, for samples that are at least half complete ($\mu = 0$, and above), there is virtually no bias. For samples that are at least 1/4 complete ($\mu = -1.2$ and above), the bias is not substantial in the censored model (though it is so in the truncated dependent variable case, as we will see in the next section).

Letting the variance of the dependent variable be estimated, which is the more realistic case, the results are somewhat different (more pessimistic). Although the asymptotic bias largely disappears as the sample becomes half or more complete, the biases are much larger in this case than the unrealistic one in which variance was known. That is, the cost of misspecification of the distribution of the error terms is higher when we do not know the variance.

It is worth noting that in both cases (variance known and unknown), the bias of the normal MLE's - which allows for correction of censoring- is much less than the bias of the sample mean. Thus it is better (less costly) to correct the mean than not to correct, even though the correction is biased. This result does not totally agree with the results of Arabmazar and Schmidt (1982). In their study, when the variance was unknown, the uncorrected sample mean was sometimes less biased than the normal MLE.

The same kind of conclusions can be derived for the bias of the estimated variance; that is, its bias depends upon the degree of censoring. As the sample becomes half or more complete there is virtually no bias at all.

2.3 The Truncated Dependent Variable Model

We now turn to the case in which we have no information concerning the unobserved observations (limit observations); even the number of limit observations is not known. The model is still as defined in (2.1), however, we observe $y_i = y_i^*$ if and only if y_i^* is non-negative. Again as was the case in the last section we assume we have only one regressor, a constant term, meaning we estimate the mean of the observations only.

The question again is: what is the statistical cost of assuming normality of the error terms, when they are in fact distributed as first-order Sargan? The numerical results are tabulated in Table 2.1 under the heading "truncated". The entries are the solutions to equation (11) of Arabmazar and Schmidt, needing only the already derived truncated moments of u given above in (2.2) and (2.3). The absolute bias is slightly greater than they found for the t_{10} distribution, in both cases of known and unknown variance. It is considerably greater than in the censored case, although it still goes to zero (very slowly) as the degree of truncation goes to zero. In the unrealistic case of variance known, for example, for samples which are 75 percent complete ($\mu = 1.2$ and above) almost no bias exists. But as the sample becomes less complete (degree of truncation rises) the bias becomes greater and greater. Going to the more realistic case of variance unknown, the bias becomes much larger, especially for heavily truncated samples. Contrary to the censored case, however, when the

variance is not known the bias is much larger than the bias of the uncorrected sample mean. (That is, it does not pay to correct for truncation.)

The robustness of the normal MLE to non-normality has also been considered by Goldberger (1980), in the truncated case, under the assumption that the only regressor is a constant term, and that the disturbance variance is known. He assumes that the true distribution of errors is a non-normal symmetric distribution (Student's t, Laplace, logistic), derives and calculates the asymptotic bias (inconsistency) of the maximum likelihood estimators which assumes the distribution of the errors to be normal. His results show that although the bias is large for largely censored samples, it goes to zero as the sample becomes complete.

2.4 The Binary (Probit) Model

In this section we take up the binary dependent variable model, which can be seen as a special case of the Tobit model, (see Goldberger (1980)). We only observe two different values for y_i (say, zero and one), according to the rule;

$$(2.4) \quad y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* < 0 \end{cases} \quad i=1,\ldots,n$$

where, y_i^* are defined in (2.1). Thus only the sign of the dependent

variable is observable. The only thing that can be identified is the ratio of the mean and standard deviation (μ / σ), so that if we assume that the variance is known and equal to one (or use σ as a normalization factor), then the mean of the dependent variable (μ) can be estimated.

The probability limit of the estimated mean (assuming σ^2 known and equal to one) is given by (12) of Arabmazar and Schmidt. The asymptotic bias is given in Table 2.1, under the heading "binary", for various values of μ .

Unlike the censored and truncated cases, the sample does not become completely observed as μ gets positive and large. The bias is smallest when $\mu = 0$, and increases (symmetrically) as μ moves in either direction from zero. The size of the bias is similar to that of the censored case with σ^2 known, for $\mu \leq 0$.

2.5 Conclusions

In this chapter we considered the inconsistency (asymptotic bias) of the normal maximum likelihood estimators for the censored, truncated and binary dependent variable models, where the true distribution of the error terms is first-order Sargan. In the case of linear regression with a fully observable dependent variable no bias exists. The normal MLE's (OLS) are consistent (they are in fact BLUE), even if normal is not the true distribution of the errors. In other words, the normal MLE's are robust to non-normality.

The results for the censored and truncated models are to some extent similar to each other and also similar to the results for the other distributions considered by Arabmazar and Schmidt (1982). First of all, the asymptotic bias disappears as μ gets large (i.e. as sample becomes complete) in both the censored and truncated models, but not in the binary case. When σ is known, the bias for all values of μ is much smaller than when we consider the more realistic case of σ unknown. This is true especially in the truncated model with a large degree of truncation.

Secondly, the bias of μ is generally less in the censored model than in the truncated model, implying that knowing the number of limit (unobserved) observations helps in getting a less biased estimate of the mean. This is so, especially when μ is very small; that is, where the sample is largely censored.

The final conclusion derived from these results is that the maximum likelihood estimator of μ (i.e., the estimator which corrects for censoring or truncation) is usually less biased than the sample mean ($\bar{y}$ or $\bar{y}^*$), suggesting that it pays to correct for censoring or truncation, even if the true error distribution is not normal.

We have only considered a very special and simple case in which there is only one regressor – a constant term – so that we are estimating only the mean of the dependent variable. Thus it is not known how different the results would be in a more general case with more than one regressor. Certainly this would be an interesting extension of this research.

Table 2.1

Asymptotic Bias of Tobit When True Errors
are First-Order Sargan

μ	$P(y^*>0)$	Sample Mean		$\mu(\sigma$ known$)$			$\mu(\sigma$ unknown$)$		σ	
		Censored	Truncated	Censored	Truncated	Binary	Censored	Truncated	Censored	Truncated
-2.8	.01	2.80	3.37	0.37	1.99	.34	-1.50	-5.53	0.75	1.31
-2.4	.01	2.41	2.97	0.23	1.63	.20	-1.23	-5.93	0.65	1.33
-2.0	.03	2.02	2.58	0.12	1.27	.08	-0.98	-6.31	0.55	1.35
-1.6	.05	1.63	2.20	0.02	0.92	-.02	-0.74	-6.84	0.44	1.37
-1.2	.10	1.26	1.81	-0.04	0.59	-.08	-0.52	-6.85	0.33	1.38
-0.8	.18	0.92	1.44	-0.08	0.29	-.11	-0.32	-6.69	0.22	1.36
-0.4	.31	0.61	1.08	-0.07	0.03	-.08	-0.15	-5.26	0.11	1.17
0	.50	0.38	0.75	-0.03	-0.14	0.0	-0.04	-2.58	0.01	0.71
0.4	.69	0.21	0.49	0.00	-0.15	.08	0.01	-0.79	-0.04	0.28
0.8	.82	0.12	0.32	0.01	-0.09	.10	0.02	-0.18	-0.04	0.07
1.2	.90	0.06	0.20	0.01	-0.03	.08	0.02	-0.01	-0.04	-0.02
1.6	.95	0.03	0.12	0.01	0.01	.02	0.01	0.03	-0.03	-0.04
2.0	.97	0.02	0.07	0.01	0.02	-.08	0.01	0.03	-0.02	-0.05
2.4	.99	0.01	0.04	0.00	0.02	-.20	0.00	0.02	-0.01	-0.04
2.8	.99	0.00	0.02	0.00	0.02	-.34	0.00	0.02	-0.01	-0.03

Chapter 3

Robustness of Sargan MLE's to Normal Errors: First Order Case

3.1 Introduction

One of the basic assumptions of the linear regression model as well as many other models is the normality of the error terms. But since in a large number of models (e.g., models with censored or limited dependent variables) the likelihood function contains the c.d.f. of the distribution of the error terms, the assumption of normality implies that the normal c.d.f. will appear in the likelihood function. The normal c.d.f. can only be evaluated numerically; the calculations would be simpler if a distribution were assumed whose c.d.f. can be expressed analytically. As we have seen in chapter 1, families of Sargan distributions are possible substitutes for the normal, as Goldfeld and Quandt (1981) have suggested.

In the previous chapter we asked the question of how costly it is to assume normality of the error terms, when they are in fact first order Sargan. In the next two chapters we ask the reverse question, which we think is more interesting. That is, what happens if the true error distribution is normal, but first or second order Sargan is mistakenly assumed[1]? In other words, how good an approximation to the normal distribution is a Sargan distribution? We will answer this question by providing some evidence on the relevent statistical

23

questions, namely the effect on one's estimates or inferences of the use of such an approximation to the (hypothesized) true error distribution.

The statistical cost involved is very much dependent upon the nature of the model in question. The misspecification of the errors' distribution (e.g., an incorrect assumption of normality) does not cause bias or inconsistency of the parameter estimates in the linear regression model, but it does cause inconsistency in limited dependent variable models, such as the Tobit model. Missiakoulis (1983) has calculated the asymptotic bias (inconsistency) for the binary (probit) model. We consider three additional models; the linear regression model, the censored regression (Tobit) model, and the truncated regression model. Although these results can not be generalized, because of the restricted set of models that has been used so far, our results serve to indicate at least to which extent the results are model dependent.

In the next three sections we will present results for three different models, but for the first-order Sargan distribution only. In the next chapter we ask the same question for the second-order Sargan distribution (for all three models). In each section, first the model and its estimators are introduced, and then the numerical calculations follow (tabulated in Table 3.1).

Most of our conclusions are similar to the conclusions of the previous chapter. For example, the bias depends stongly on the degree of censoring or truncation, and is generally larger when σ^2 is unknown than when it is known. However, an interesting result is that it is

typically less costly (in terms of asymptotic bias) to assume Sargan errors when the truth is normal than it is to assume normal errors when the truth is Sargan.

3.2 Linear Regression Model

The model considered in this section is the linear regression model;

$$y_i = \sum_{j=1}^{K} X_{ij} \beta_j + u_i$$

(3.1)   or

$$y_i = \beta' X_i + u_i , \qquad i = 1, 2, \ldots, n .$$

where, X_i represents a row vector of the regressors, with K elements, and β is a column vector of coefficients with K elements. The error terms, u_i, are i.i.d with zero mean and are independent of the regressors. What remains to be discussed is the distribution of the errors.

In section 2.2 we asked the question previously discussed by Goldfeld and Quandt (1981), namely, what happens if we assume normal errors when in fact the true distribution is first order Sargan. It is obvious that normal MLE will simply be OLS , and that OLS is BLUE, unbiased and consistent, under above conditions. However, Goldfeld and Quandt provided the non-trivial result that the efficiency of OLS relative to the Sargan maximum likelihood estimator is approximately 0.84 .

In this section, the opposite question is addressed. Thus the errors are actually normal, but we treat them as first-order

Sargan. (We will ask the same question with the second-order Sargan instead in the next chapter.) We therefore discuss the cost of using first-order Sargan as an approximation for the normal distribution, when the normal distribution is the correct one. The result, as we will show, is that there is no cost (no asymptotic bias) involved at all in assuming first-order Sargan when normal is the true error distribution. Our result indeed is much stronger than just stated, since it says that there is no bias if first-order Sargan is assumed mistakenly as long as the true distribution of the error terms is symmetric around zero. That is, the result hinges on the symmetry of the true distribution of the error terms only and not on its normality. Therefore, Sargan "MLE" is consistent provided that error terms are i.i.d with a symmetric distribution around zero.[2]

In order to show this result, we begin by forming the Sargan likelihood function.

$$(3.2) \quad \ln L = n \ln \alpha - n \ln 4 - \alpha \sum_i |y_i - \beta'X_i| + \sum_i \ln [1+\alpha|y_i - \beta'X_i|].$$

The Sargan MLE's ($\hat{\alpha}$, $\hat{\beta}$) satisfy the following first order conditions:

$$(3.3) \quad \frac{\partial \ln L}{\hat{\alpha}} = \frac{n}{\hat{\alpha}} - \sum_i |y_i - \hat{\beta}'X_i| + \sum_i \frac{|y_i - \hat{\beta}'X_i|}{1+\hat{\alpha}|y_i - \hat{\beta}'X_i|} = 0$$

$$(3.4) \quad \frac{\partial \ln L}{\partial \hat{\beta}_j} = \hat{\alpha} \sum_{+} X_{ij} - \hat{\alpha} \sum_{-} X_{ij} - \hat{\alpha} \sum_{+} \frac{1}{1+\hat{\alpha}(y_i - \hat{\beta}'X_i)} X_{ij}$$

$$+ \hat{\alpha} \sum_{-} \frac{1}{1-\hat{\alpha}(y_i - \hat{\beta}'X_i)} X_{ij} = 0$$

which are similar to the equations (3.2) in page 146 of Goldfeld and Quandt. Here $\sum_{+}$ refers to summation over observations such that: $y_i \geq \hat{\beta}'X_i$, and conversely for $\sum_{-}$.

Now we divide both (3.3) and (3.4) by n (sample size) and take probability limits. Defining $\tilde{\alpha}$ and $\tilde{\beta}$ to be the probability limits of $\hat{\alpha}$ and $\hat{\beta}$ respectively, we get;

$$(3.5) \quad \frac{1}{\tilde{\alpha}} - \text{plim} \frac{1}{n} \sum_i |y_i - \tilde{\beta}'X_i| + \text{plim} \frac{1}{n} \sum_i \frac{|y_i - \tilde{\beta}'X_i|}{1+\tilde{\alpha}|y_i - \tilde{\beta}'X_i|} = 0$$

$$(3.6) \quad \tilde{\alpha} \, \text{plim} \frac{n_+}{n} \, \text{plim} \frac{1}{n_+} \sum_{+} X_{ij} - \tilde{\alpha} \, \text{plim} \frac{n_-}{n} \, \text{plim} \frac{1}{n_-} \sum_{-} X_{ij}$$

$$- \tilde{\alpha} \, \text{plim} \frac{n_+}{n} \, \text{plim} \frac{1}{n_+} \sum_{+} \frac{1}{1+\tilde{\alpha}(y_i - \tilde{\beta}'X_i)} X_{ij}$$

$$+ \tilde{\alpha} \, \text{plim} \frac{n_-}{n} \, \text{plim} \frac{1}{n_-} \sum_{-} \frac{1}{1-\tilde{\alpha}(y_i - \tilde{\beta}'X_i)} X_{ij} = 0.$$

(Here n_+ is the number of terms in $\sum_{+}$, and similary for n_-).

Deriving some of the probability limits in the above two equations is not feasible. But a closer look at these equations will indicate that to show consistency of $\tilde{\beta}$, we do not need to evaluate all of these probability limits explicitly. Rather we will simply show

that equations (3.5) and (3.6) are satisfied with $\tilde{\beta}_j = \beta_j$, for some $\tilde{a}$.

Starting with (3.5), we set $\tilde{\beta}_j = \beta_j$ for all $j = 1,2,\ldots,K$;

$$(3.7) \quad \frac{1}{\tilde{\alpha}} - \text{Plim} \frac{1}{n} \sum_i |u_i| + \text{Plim} \frac{1}{n} \sum_i \frac{|u_i|}{1+\tilde{\alpha}|u_i|} = 0$$

or

$$(3.8) \quad \frac{1}{\tilde{\alpha}} - E|u| + E \left[\frac{|u|}{1+\tilde{\alpha}|u|}\right] = 0 .$$

If u is distributed normally with mean zero and variance σ^2, the first expected value in (3.8) is easy to evaluate ($\sigma \sqrt{2/\pi}$), but the second expected value poses a problem. However, as long as there exists a positive solution $\tilde{a}$ to (3.8), its actual value is not of any importance.

Next we define;

$$(3.9) \quad \tilde{X}_j = \text{plim} \frac{1}{n} \sum_i X_{ij} = \text{plim} \frac{1}{n_+} \sum_+ X_{ij} = \text{plim} \frac{1}{n_-} \sum_- X_{ij} ,$$

where the last two equations follow from the the fact that X is independent of u.

With $\tilde{\beta}_j = \beta_j$ (for all $j=1,2,\ldots,K$), equation (3.6) can be rewritten as

$$(3.10) \quad \tilde{\alpha} \text{ Plim} \frac{n_+}{n} \tilde{X}_j - \tilde{\alpha} \text{ Plim} \frac{n_-}{n} \tilde{X}_j - \tilde{\alpha} \text{ Plim} \frac{n_+}{n} \text{Plim} \frac{1}{n_+} \sum_+ \frac{1}{1+\tilde{\alpha}u_i} X_{ij}$$

$$+ \tilde{\alpha} \text{ Plim} \frac{n_-}{n} \text{Plim} \frac{1}{n_-} \sum_- \frac{1}{1-\tilde{\alpha}u_i} X_{ij} = 0$$

$$, j = 1, \ldots, K$$

Also note that;

$$(3.11) \quad \text{Plim} \frac{n_+}{n} = P(y_i > \tilde{\beta}'X_i) = P(u_i > \tilde{\beta}'X_i - \beta'X_i)$$

$$(3.12) \quad \text{Plim} \frac{n_-}{n} = P(y_i < \tilde{\beta}'X_i) = P(u_i < \tilde{\beta}'X_i - \beta'X_i)$$

and, with $\tilde{\beta}_j = \beta_j$ (for all $j=1,2,\ldots,K$) both expressions in (3.11) and (3.12) are equal to 1/2 with any symmetric distribution of the error terms. Also,

$$(3.13) \quad \text{plim} \frac{1}{n_+} \sum_+ [\frac{1}{1+\tilde{\alpha}u_i} X_{ij}] = E[\frac{1}{1+\tilde{\alpha}u} X_j | u > 0]$$

$$= E[\frac{1}{1+\tilde{\alpha}u} | u > 0] \tilde{X}_j ,$$

and similary for the similar term involving n_-, again because X and u are independent. Then (3.10) becomes;

$$(3.14) \quad \tilde{\alpha} P(u > 0) \tilde{X}_j - \tilde{\alpha} P(u < 0) \tilde{X}_j - \tilde{\alpha} P(u > 0) E[\frac{1}{1+\tilde{\alpha}u} | u > 0] \tilde{X}_j$$

$$+ \tilde{\alpha} P(u < 0) E[\frac{1}{1-\tilde{\alpha}u} | u < 0] \tilde{X}_j = 0$$

It is obvious that (3.14) holds, provided that u is distributed symmetrically around zero (of which the normal distribution is a special case) and also provided that the necessary expected values exist. This is so because symmetry implies:

$$P(u \geq 0) = P (u \leq 0) , \quad \text{and}$$

$$(3.15) \quad E \left[\frac{1}{1+\tilde{\alpha} u} \mid u > 0 \right] = E \left[\frac{1}{1-\tilde{\alpha} u} \mid u < 0 \right]$$

Therefore, both equations (3.5) and (3.6) are indeed satisfied for $\tilde{\beta} = \beta$, and for some value of $\tilde{\alpha}$ as implied by (3.8).

This result indicates that the Sargan MLE's of the coefficients of the linear regression model are asymptotically unbiased (consistent), and thus there is no loss (in terms of asymptotic bias) if any symmetric distribution of the error terms is approximated by first-order Sargan. (The same result holds for second order Sargan, as is shown in the next chapter.) Therefore, the result indicates a sense in which the Sargan distribution is an adequate approximation for any symmetric distribution of error terms in the linear regression model.

3.3 The Tobit Model

We now turn to the censored dependent variable model, in which we assume that there is only one regressor, a constant term. This assumption is made for reasons of tractability, and also to make the analysis comparable to the studies done by Goldberger (1981), Arabmazar and Schmidt (1982), and Missiakoulis (1983). Our attempt, therefore, is to estimate the mean of the original (uncensored) population. This can be seen as a starting point for a more complicated case in which we have several regressors in the regression.

Thus we consider the model:

$$(3.16) \quad y_i^* = \mu + u_i \qquad i = 1, 2, \ldots, n$$

where u_i are i.i.d with zero mean. However, we do not observe y_i^*, but only $y_i = \max (0, y_i^*)$. That is, we have a model with a censored dependent variable; it would be the Tobit model if the u_i were distributed as normal.

In section 2.3 we asked the question of what happens if the error terms actually have a first order Sargan distribution, but they are assumed to be normal. However, here we ask the opposite question. That is; if the errors are actually normal but we treat them as first order Sargan, what happens to the estimates of the mean? In other words, how good is the first-order Sargan distribution as an approximation to the normal distribution? Note that again we address the cost involved (in approximation) only in terms of the asymptotic bias of the estimators. Basically we expect the same kind of results as in section 2.3, meaning that the bias, although possibly large in cases of severe censoring, will disappear as we get closer to a complete sample.

In order to measure this inconsistency, we start by looking at Sargan logarithm of likelihood function ;

$$(3.17) \quad \ln L = m \ln \alpha - m \ln 4 - \alpha \sum_{i=1}^{m} |y_i - \mu| + \sum_{i=1}^{m} \ln [1 + \alpha |y_i - \mu|]$$

$$+ (n-m) \ln F(-\mu) ,$$

where, we have assumed that the first m observations $(y_1, y_2, \ldots, y_m)$ are positive, and the remaining $(n-m)$ observations $(y_{m+1}, y_{m+2}, \ldots, y_n)$ are all zero (limit observations). F is the first-order Sargan c.d.f., defined as (a special case of (1.9));

$$(3.18) \quad F(x) = \begin{cases} 1 - \frac{1}{4} e^{-\alpha x} (2+\alpha x) & x > 0 \\[3mm] \frac{1}{4} e^{\alpha x} (2-\alpha x) & x < 0 \end{cases}$$

The first order conditions for the α and μ are as follows;

$$(3.19) \quad \frac{\partial \ln L}{\partial \hat{\alpha}} = \frac{m}{\hat{\alpha}} - \sum_{i=1}^{m} |y_i - \hat{\mu}| + \sum_{i=1}^{m} \left[\frac{|y_i - \hat{\mu}|}{1 + \hat{\alpha}|y_i - \hat{\mu}|} \right]$$
$$+ (n-m)\, G\, (\hat{\alpha}, \hat{\mu}) = 0$$

$$(3.20) \quad \frac{\partial \ln L}{\partial \hat{\mu}} = \hat{\alpha}(m_+ - m_-) - \hat{\alpha} \sum_+ \frac{1}{1 + \hat{\alpha}(y_i - \hat{\mu})} + \hat{\alpha} \sum_- \frac{1}{1 - \hat{\alpha}(y_i - \hat{\mu})}$$
$$- (n-m)\, H\, (-\hat{\mu}) = 0$$

Here G is defined as:

$$(3.21) \quad G(\alpha, \mu) = \begin{cases} -\mu + \mu/(2+\alpha\mu) & \mu > 0 \\[3mm] \dfrac{e^{\alpha\mu} (\alpha\mu^2 - \mu)}{4 - e^{\alpha\mu} (2-\alpha\mu)} & \mu < 0 \end{cases}$$

and H is defined as the ratio of the Sargan density function f to the Sargan c.d.f., with $\gamma_0 = \gamma_1 = 1$ (in order to insure the uniqueness of the

MLE in this case).

$$(3.22) \quad H(-x) = \frac{f(-x)}{F(-x)} = \begin{cases} \dfrac{\alpha(1+\alpha x)}{2+\alpha x} & x > 0 \\[20pt] \dfrac{\alpha e^{\alpha x}(1-\alpha x)}{4-e^{\alpha x}(2-\alpha x)} & x < 0 \end{cases}$$

Also, $\overset{m}{\underset{+}{\sum}}$ refers to summation over observations such that both conditions; $y \geq 0$, $y \leq \hat{\mu}$ hold, while $\overset{m}{\underset{-}{\sum}}$ refers to summation over observations such that both $y \geq 0$, $y \leq \hat{\mu}$ hold, and m_{+}, m_{-} represnt the relevent numbers of observations.

Now dividing by m and taking probability limits, with $\tilde{\mu} = \text{Plim } \hat{\mu}$, $\tilde{\alpha} = \text{Plim } \hat{\alpha}$, we get:

$$(3.23) \quad \frac{1}{\tilde{\alpha}} - \text{Plim} \frac{1}{m} \sum_{i=1}^{m} |y_i - \tilde{\mu}| + \text{Plim} \frac{1}{m} \sum_{i=1}^{m} \frac{|y_i - \tilde{\mu}|}{1+\tilde{\alpha}|y_i - \tilde{\mu}|}$$

$$+ \text{Plim} \frac{n-m}{m} G(\tilde{\alpha}, \tilde{\mu}) = 0$$

$$(3.24) \quad \tilde{\alpha} \left[\text{plim} \frac{m+}{m} - \text{plim} \frac{m-}{m} \right] - \tilde{\alpha} \, \text{plim} \frac{m+}{m} \, \text{plim} \frac{1}{m_{+}} \sum_{+} \frac{1}{1+\tilde{\alpha}(y_i - \tilde{\mu})}$$

$$+ \tilde{\alpha} \, \text{plim} \frac{m-}{m} \, \text{plim} \frac{1}{m_{-}} \sum_{-} \frac{1}{1-\tilde{\alpha}(y_i - \tilde{\mu})} - \text{plim} \frac{n-m}{m} H(-\tilde{\mu}) = 0$$

The probability limits can be rewritten as:

$$(3.25) \quad \frac{1}{\tilde{\alpha}} - E[|y - \tilde{\mu}| \mid y > 0] + E\left[\frac{|y - \tilde{\mu}|}{1+\tilde{\alpha}|y - \tilde{\mu}|} \mid y > 0\right] + \frac{P(y < 0)}{P(y > 0)} G(\tilde{\alpha}, \tilde{\mu}) = 0 \, .$$

$$\tilde{\alpha}[P(y>\tilde{\mu}|y>0) - P(y<\tilde{\mu}|y>0)]$$

$$-\tilde{\alpha}\, P(y>\tilde{\mu}|y>0)\ E[\frac{1}{1+\tilde{\alpha}(y-\tilde{\mu})} \mid y>\tilde{\mu},\ y>0]$$

$$(3.26)$$

$$+\tilde{\alpha}\, P(y<\tilde{\mu}|y>0)\ E[\frac{1}{1-\tilde{\alpha}(y-\tilde{\mu})} \mid y<\tilde{\mu},\ y>0]$$

$$-\frac{P(y<0)}{P(y>0)} H(-\tilde{\mu}) = 0$$

The probability limits $\tilde{\alpha}$, $\tilde{\mu}$ of the Sargan "MLE's" will satisfy these equations. To evaluate the inconsistency that results when the error terms are actually normal (rather than Sargan), we need to evaluate (3.23) and (3.24) for Y's normally distributed with mean μ and variance σ^2. There is no problem in evaluating the required probabilities, nor in calculating $E\ [\ |Y - \tilde{\mu}\ ||\ Y \geq 0\ ]$ analytically (see appendix A). However, the other three expected values in (3.23) and (3.24) were intractable and had to be calculated by Monte Carlo methods (i.e., simulated). The number of drawings (of y) which we used ranged from 10,000 to 130,000 depending upon the likelihood of the conditioning event - for example, if $\mu = -1$; $\tilde{\mu}$ =3 it will take a lot of drawings to observe $y \geq \tilde{\mu}$ enough times to calculate the conditional mean accurately. We also calculated some of the more difficult expected values by numerical integration, and got essentially the same results, The equations (3.23) and (3.24) were then solved using the Gauss-Newton method.[3]

Our results are given in Table 3.1, under the heading "censored". (The results under "truncated" will be discussed in the

next section.) We first look at the case where the <u>true</u> a is known (that is, the <u>true</u> a is taken to equal two, which corresponds to $\sigma^2 = 1$.) and then consider the case where the variance is not known, where the latter case is presumably the realistic one. The results are qualitatively similar to those of Table 2.1, for the opposite case (Sargan is true but normal is assumed), especially in the rapid disappearance of the asymptotic bias as the degree of censoring falls. That is, the bias is heavily dependent on how complete the samples are. For samples that are at least half complete, there is virtually no bias, and for the samples that are at least 1/4 complete ($\mu = -1.2$ and above), the asymptotic bias is small. (However, as we will see in the next section, this is not so in the truncated case, in which the bias is generally larger than in the censored case.)

Now we relax the assumption that the variance is known, to see how much of a difference it makes whether the variance is known. The results, as shown in Table 3.1, clearly show that knowing the variance does matter. The bias generally is larger when we estimate the variance of the dependent variable (σ^2), than when we assume it. That is, the results for the case in which the variance is unknown are more pessimistic than in the case of known varaince. Indeed, the difference this makes is substantial, at least for heavily censored samples; the absolute bias is typically two or more times higher when the variance is unknown than when it is known, for samples less than 1/4 complete.

Another interesting result is that the bias is typically smaller in Table 3.1 than in Table 2.1. It is less costly (in terms of absolute asymptotic bias) to assume Sargan when the truth is normal

than it is to assume normal when the truth is Sargan. Why this should be so is not obvious. (Note that it is not so when the variance is assumed known.)

In the next chapter, we ask the same question again, for the second-order Sargan distribution. That is, we will measure the statistical cost of using second order Sargan as an approximation to normal. Not surprisingly, there is less asymptotic bias involved when we use second-order Sargan than first-order. This may indicate that the second order Sargan distribution is a more adequate approximation to the normal in the econometric models. However, we have selected quite a limited number of models to investigate the cost, so the generality of this conclusion is perhaps questionable.

3.4 The Truncated Dependent Variable Model

In this section we discuss the truncated version of the Tobit model of the last section. That is, the model is given by (3.16). However, now we observe $y_i = y_i^*$ if and only if $y_i^* \geq 0$. Nothing else is observed, not even the number of unobserved observations (limit observations). The results calculated in this section are in many ways similar to those for the censored dependent variable model of the last section. The bias is strongly dependent upon the degree of truncation, going to zero as sample becomes complete, and it is also larger when the error variance is unknown. However, the bias is generally larger in the truncated case than in the censored case.

We wish to evaluate the asymptotic bias due to assuming the

errors to be first-order Sargan when they are in fact normal. We start
with the Sargan likelihood function:

$$(3.27) \quad \ln L = -m \ln 4 + m \ln \alpha - \alpha \sum_{i=1}^{m} |y_i - \mu| + \sum_{i=1}^{m} \ln[1 + \alpha|y_i - \mu|]$$
$$- m \ln F(\mu) .$$

As in the previous section, we first find the first order
conditions:

$$(3.28) \quad \frac{\partial \ln L}{\partial \hat{\alpha}} = \frac{m}{\hat{\alpha}} - \sum_{i=1}^{m} |y_i - \hat{\mu}| + \sum_{i=1}^{m} \frac{|y_i - \hat{\mu}|}{1 + \hat{\alpha}|y_i - \hat{\mu}|} - m R(\hat{\alpha}, \hat{\mu}) = 0$$

$$(3.29) \quad \frac{\partial \ln L}{\partial \hat{\mu}} = \hat{\alpha}(m_+ - m_-) - \hat{\alpha} \sum \frac{1}{1 + \hat{\alpha}(y_i - \hat{\mu})} + \hat{\alpha} \sum \frac{1}{1 - \hat{\alpha}(y_i - \hat{\mu})}$$
$$- m V(\hat{\mu}) = 0$$

where,

$$(3.30) \quad R(\alpha, \mu) = \begin{cases} \dfrac{\mu(1 + \alpha\mu)e^{-\alpha\mu}}{4 - e^{-\alpha\mu}(2 + \alpha\mu)} & \mu > 0 \\[2em] \dfrac{\mu(1 - \alpha\mu)}{2 - \alpha\mu} & \mu < 0 \end{cases}$$

and

$$(3.31) \quad V(\mu) = \frac{f(\mu)}{F(\mu)} = \begin{cases} \dfrac{\alpha(1 + \alpha\mu)e^{-\alpha\mu}}{4 - (2 + \alpha\mu)e^{-\alpha\mu}} & \mu > 0 \\[2em] \dfrac{\alpha(1 - \alpha\mu)}{2 - \alpha\mu} & \mu < 0 \end{cases}$$

and other notation is as used in the previous section.

Then, dividing by m (sample size) and taking probability
limits yields the following equations, which the probability limits

$\tilde{a}$, $\tilde{\mu}$ will satisfy:

$$(3.32) \quad \frac{1}{\tilde{\alpha}} - E[|y-\tilde{\mu}| \,\Big|\, y>0] + E[\frac{|y-\tilde{\mu}|}{1+\tilde{\alpha}|y-\tilde{\mu}|} \,\Big|\, y>0] - R(\tilde{\alpha},\tilde{\mu}) = 0$$

$$(3.33) \quad \tilde{\alpha}P(y>\tilde{\mu}|y>0) - \tilde{\alpha}P(y<\tilde{\mu}|y>0) - \tilde{\alpha}P(y>\tilde{\mu}|y>0)\, E[\frac{1}{1+\tilde{\alpha}(y-\tilde{\mu})} \,\Big|\, y>\tilde{\mu},y>0]$$

$$+ \tilde{\alpha}P(y<\tilde{\mu}|y>0)\, E[\frac{1}{1-\tilde{\alpha}(y-\tilde{\mu})} \,\Big|\, y<\tilde{\mu},y>0]$$

$$- V(\tilde{\mu}) = 0$$

Again, since the normal c.d.f will appear in the expected values in equations (3.32) and (3.33), as in the last section, we are not able to calculate them analytically (except the first one). Therefore, we solve (3.32) and (3.33) numerically for $\tilde{a}$ and $\tilde{\mu}$, with the same simulation procedure used in the previous section used again to evaluate the necessary expected values. Our results are given in Table 3.1, under the heading "truncated".

We will discuss both the case of known a and the case of unknown a . Generally speaking, in both cases the bias is much larger than in the corresponding censored model. In the truncated model, the asymptotic bias goes to zero (in both cases) as samples become about 75% complete, compare this to the censored model in which the bias is virtually zero when samples are only half complete. As in the censored model, the results are heavily dependent upon the degree of truncation. For low degrees of truncation, the bias is very small, while at higher degrees of truncation the bias gets very large. Basically the same kind of conclusions can be drawn for the asymptotic bias of a , for the unknown variance case, of course. That is, the bias is larger in the

truncated model than in the censored model, and in both cases, the bias of a depends on the degree of truncation (or censoring).

Consider first the (realistic) case in which a is treated as unknown. The absolute bias is clearly greater than for the corresponding censored case, as just noted. However, what is most striking is how much smaller the absolute bias is than in the opposite case (Table 2.1), in which we assumed that first-order Sargan was the true distribution of the error terms, but by mistake we used normal instead. The same kind of results were also found in the censored model, though not so strongly. It is less costly (in the truncated model, far less costly) to assume Sargan when the truth is normal than it is to assume normal when the truth is Sargan. Nevertheless, the bias is still substantial except for samples that are substantially untruncated.

Next we consider the (unrealistic) case in which a is assumed to be known. In the present case (normal errors with zero mean and variance 1) this means we set $\tilde{a} = 2$, and then solve (3.33) for $\tilde{\mu}$. This presents some computational problems, since for any fixed $\tilde{a}$, the left hand side of (3.33) approaches zero as $\tilde{\mu}$ approaches $-\infty$. In other words, $\tilde{\mu} = -\infty$ is always a solution. This is not a problem in and of itself as long as another solution exists. However, for $\tilde{\mu} \leq -0.8$, apparently no other solution existed. (At least we could not find one, despite our best efforts.) The corresponding entries in Table 3.1 are marked "not available". Those biases which are available ($\mu \geq -.8$) are reasonable, however.

3.5 Conclusions

In this chapter, we attempt to measure the statistical cost of the use of the first-order Sargan distribution as an approximation to the normal, in econometric models. Since the motivation for the Sargan distribution is really computational ease, it seems reasonable to focus on the case in which the errors are actually normal, but the Sargan "MLE" is used instead. Our measure of the statistical cost of such an approximation is the inconsistency (asymptotic bias) which results.

An obvious feature of this "cost" is that it is model dependent. No asymptotic bias results in the linear regression model. As we saw in section 3.2, the unbiaseness of the coefficients hinges only on the symmetry of the distribution of the errors and not on its normality. In the censored regression (Tobit) model, the bias can be substantial, but it is reasonably small for samples that are at least 50% complete. In the truncated regression model, the bias is even larger, and seems large enough to be bothersome except in samples that are substantially (say, 75%) complete.

Also the results indicate that knowing the error variance helps, in both the censored and truncated models. The bias is much smaller when we know a than when a is estimated.

Finally, comparing these results to those of chapter 2 tells us that the Sargan distribution is a better approximation for the normal than the normal is for the Sargan. That is, the bias is much smaller when the world is normal but we approximate it by Sargan, than when the world is Sargan, but we treat it as normal.

Obviously other models could be considered, (e.g., the disequilibrium model, which Goldfeld and Quandt (1981) used to introduce these distributions), but the fact that the adequacy of the approximation is so clearly model dependent is sufficient to argue for caution in the use of these distributions. At least this is so if they are really viewed as approximations to the normal; it is possible to argue that there is no more reason to assume normal errors than to assume Sargan errors.

In this chapter we only considered first-order Sargan distribution, since higher order Sargan distributions contain more parameters than the normal. It is the case that a higher order Sargan distribution (e.g., second-order) would perform more adequately, as we will see in the following chapter.

Footnotes

1) This is perhaps a strange case to consider, since Sargan errors are unlikely to be assumed in a regression context, being <u>less</u> convenient than normal errors in this case. However, the regression case serves as a useful standard of comparison for the censored and truncated cases to be considered later.

2) Subject to the qualification that certain moments need to exist, as will be made explicit below.

3) Note that this is an expensive undertaking because $\tilde{a}$ and $\tilde{\mu}$ change at each iteration, thus requiring fresh calculation of the expected values just discussed at each iteration.

Table 3.1

Asymptotic Bias of Sargan "MLE" When True Errors are N(0,1)

μ	P(y[*]>0)	μ(α known)		μ(α unknown)		α	
		Censored	Truncated	Censored	Truncated	Censored	Truncated
-2.8	.003	-0.58	*	0.97	2.81	1.70	2.97
-2.4	.008	-0.32	*	0.77	2.30	1.32	2.18
-2.0	.023	-0.12	*	0.59	1.97	1.00	1.83
-1.6	.055	0.01	*	0.46	1.48	0.78	1.28
-1.2	.115	0.09	*	0.33	1.15	0.53	0.98
-0.8	.212	0.10	-0.87	0.20	0.86	0.30	0.70
-0.4	.345	0.08	0.19	0.09	0.58	0.08	0.44
0	.500	0.03	0.23	0.02	0.34	-0.08	0.20
0.4	.655	0.00	0.16	0.00	0.17	-0.15	0.04
0.8	.788	0.00	0.08	0.00	0.07	-0.16	-0.05
1.2	.885	0.00	0.03	0.00	0.00	-0.14	-0.14
1.6	.945	0.00	0.01	0.00	-0.02	-0.12	-0.16
2.0	.977	0.00	-0.01	0.00	-0.03	-0.11	-0.18
2.4	.992	0.00	-0.01	0.00	-0.02	-0.11	-0.18
2.8	.997	0.00	-0.01	0.00	-0.01	-0.10	-0.15

[*] Not available. See text for explanation.

Appendix A;

1. **if $\tilde{\mu} < 0$**

$$E\{|y-\tilde{\mu}||y>0\} = \int_0^\infty (y-\tilde{\mu}) f(y|y>0)\, dy$$

$$= \int_0^\infty yf(y|y>0)\, dy - \tilde{\mu} \int_0^\infty f(y|y>0)dy$$

$$= E(y|y>0) - \tilde{\mu} P(y>0|y>0)$$

$$E\{|y-\tilde{\mu}|y>0\} = \mu + \sigma \cdot m\left(\frac{\mu}{\sigma}\right) - \tilde{\mu} \tag{A-1}$$

where $m\left(\frac{\mu}{\sigma}\right) = \dfrac{\text{std. normal density } \left(\frac{\mu}{\sigma}\right)}{\text{std. normal cdf } \left(\frac{\mu}{\sigma}\right)}$

2. **if $\tilde{\mu} > 0$**

$$E\{|y-\tilde{\mu}||y>0\} = \int_0^\infty |y-\tilde{\mu}| f(y|y>0)\, dy$$

$$= \int_0^{\tilde{\mu}} (\tilde{\mu}-y) f(y|y>0)\, dy + \int_{\tilde{\mu}}^\infty (y-\tilde{\mu}) f(y|y>0)dy$$

$$= \tilde{\mu} \int_0^{\tilde{\mu}} f(y|y>0)\, dy - \int_0^{\tilde{\mu}} yf(y|y>0)dy$$

$$+ \int_{\tilde{\mu}}^\infty yf(y|y>0)dy - \tilde{\mu} \int_{\tilde{\mu}}^\infty f(y|y>0)dy \tag{A-2}$$

$$\int_0^{\tilde{\mu}} f(y|y>0) = \frac{1}{P(y>0)} \int_0^{\tilde{\mu}} f(y)dy = \frac{P(0<y<\tilde{\mu})}{P(y>0)} = \frac{\Phi(\frac{\tilde{\mu}-\mu}{\sigma}) - \Phi(\frac{-\mu}{\sigma})}{\Phi(\frac{M}{\sigma})}$$

$$\tag{A-3}$$

$$\int_{\tilde{\mu}}^\infty f(y|y>0) = \frac{1}{P(y>0)} \int_{\tilde{\mu}}^\infty f(y)dy = \frac{\Phi(\frac{\mu-\tilde{\mu}}{\sigma})}{\Phi(\frac{\mu}{\sigma})}$$

44

$$\int_0^{\tilde{\mu}} yf(y|y>0)dy = \frac{P(0<y<\tilde{\mu})}{P(y>0)} \int_0^{\tilde{\mu}} yf(y|0<y<\tilde{\mu})dy$$

$$= \frac{P(0<y<\tilde{\mu})}{P(y>0)} E(y|0<y<\tilde{\mu}) \tag{A-4}$$

$$= \frac{\Phi(\frac{\mu}{\sigma})-\Phi(\frac{\mu-\tilde{\mu}}{\sigma})}{\Phi(\frac{\mu}{\sigma})} \left[\mu+\sigma \frac{\phi(\frac{-\mu}{\sigma})-\phi(\frac{\tilde{\mu}-\mu}{\sigma})}{\Phi(\frac{\tilde{\mu}-\mu}{\sigma})-\Phi(\frac{-\mu}{\sigma})}\right] \tag{A-5}$$

$$\int_{\tilde{\mu}}^{\infty} yf(y|y>0)dy = \frac{P(\tilde{\mu}<y<\infty)}{P(y>0)} E[y|\tilde{\mu}<y]$$

$$= \frac{1-\Phi(\frac{\tilde{\mu}-\mu}{\sigma})}{\Phi(\frac{\mu}{\sigma})} \left[\mu+\sigma \frac{\phi(\frac{\mu-\tilde{\mu}}{\sigma})}{\Phi(\frac{\mu-\tilde{\mu}}{\sigma})}\right] \tag{A-6}$$

$$E\{|y-\tilde{\mu}| \big| y>0\} = (\mu-\tilde{\mu}) \left[\frac{2\Phi(\frac{\mu-\tilde{\mu}}{\sigma})-\Phi(\frac{\mu}{\sigma})}{\Phi(\frac{\mu}{\sigma})}\right] + \sigma \frac{2\phi(\frac{\mu-\tilde{\mu}}{\sigma})-\phi(\frac{\mu}{\sigma})}{\Phi(\frac{\mu}{\sigma})} \tag{A-7}$$

Robustness of Sargan MLE's to Normal Errors: Second Order Case

4.1 Introduction

In this section we assume a second-order Sargan density to estimate the parameters of the three models in the last chapter. First we discuss the linear regression model, then the Tobit model and finally the truncated dependent variable model.

4.2 The Linear Regression Model

We begin our discussion with the linear regression model as defined in (2.1), where the error terms are i.i.d with zero mean, and are independent of the regressors. We consider the case where the true error distribution is normal, but mistakenly it is assumed to be second-order Sargan. The results are the same as in the case of first-order Sargan, that is, if the error terms are i.i.d normal (or indeed have any distribution symmetric around zero), the Sargan MLE is consistent. To show this, we start with the log likelihood function;

$$(4.1) \quad \ln L = n \ln \alpha - n \ln 4 - n \ln (1+\gamma_2) - \alpha \sum_i | y_i - \beta' X_i |$$

$$+ \sum_i \ln [1+\alpha|y_i - \beta' X_i| + \alpha^2 \gamma_2 (y_i - \beta' X_i)^2]$$

The first order conditions are[1]:

$$(4.2) \quad \frac{n}{\hat{\alpha}} - \Sigma_i \, | \, y_i - \hat{\beta}'X_i | + \Sigma_i \frac{|y_i - \hat{\beta}'X_i| + 2\hat{\alpha}\hat{\gamma}_2(y_i - \hat{\beta}'X_i)^2}{1 + \hat{\alpha}|y_i - \hat{\beta}'X_i| + \hat{\alpha}^2\hat{\gamma}_2(y_i - \hat{\beta}'X_i)^2} = 0$$

$$(4.3) \quad \frac{-n}{1+\hat{\gamma}_2} + \Sigma_i \frac{\hat{\alpha}^2(y_i - \hat{\beta}'X_i)^2}{1 + \hat{\alpha}|y_i - \hat{\beta}'X_i| + \hat{\alpha}^2\hat{\gamma}_2(y_i - \hat{\beta}'X_i)^2} = 0$$

$$(4.4) \quad \hat{\alpha}(\Sigma_{i+} X_i - \Sigma_{i-} X_i) - \hat{\alpha} \Sigma_{+} \frac{1 + 2\hat{\alpha}\hat{\gamma}_2(y_i - \hat{\beta}'X_i)}{1 + \hat{\alpha}(y_i - \hat{\beta}'X_i) + \hat{\alpha}^2\hat{\gamma}_2(y_i - \hat{\beta}'X_i)^2} X_i$$

$$+ \hat{\alpha} \Sigma_{-} \frac{1 - 2\hat{\alpha}\hat{\gamma}_2(y_i - \hat{\beta}'X_i)}{1 - \hat{\alpha}(y_i - \hat{\beta}'X_i) + \hat{\alpha}^2\hat{\gamma}_2(y_i - \hat{\beta}'X_i)^2} X_i = 0$$

Now we divide (4.2)-(4.4) by n and take the

$$\frac{1}{\tilde{\alpha}} - \text{Plim} \frac{1}{n} \Sigma_i |y_i - \tilde{\beta}'X_i| + \text{Plim} \frac{1}{n} \Sigma_i \frac{|y_i - \tilde{\beta}'X_i| + 2\tilde{\alpha}\tilde{\gamma}_2(y_i - \tilde{\beta}'X_i)^2}{1 + \tilde{\alpha}|y_i - \tilde{\beta}'X_i| + \tilde{\alpha}^2\tilde{\gamma}_2(y_i - \tilde{\beta}'X_i)^2} = 0$$

$$(4.5)$$

$$(4.6) \quad \frac{-1}{1+\tilde{\gamma}_2} + \text{Plim} \frac{1}{n} \Sigma_i \frac{\tilde{\alpha}^2(y_i - \tilde{\beta}'X_i)^2}{1 + \tilde{\alpha}|y_i - \tilde{\beta}'X_i| + \tilde{\alpha}^2\tilde{\gamma}_2(y_i - \tilde{\beta}'X_i)^2} = 0$$

$$(4.7) \quad \tilde{\alpha} \, [\text{Plim} \frac{1}{n} \Sigma_{+} X_i - \text{Plim} \frac{1}{n} \Sigma_{-} X_i]$$

$$- \tilde{\alpha} \, \text{Plim} \frac{1}{n} \Sigma_{+} \frac{1 + 2\tilde{\alpha}\tilde{\gamma}_2(y_i - \tilde{\beta}'X_i)}{1 + \tilde{\alpha}(y_i - \tilde{\beta}'X_i) + \tilde{\alpha}^2\tilde{\gamma}_2(y_i - \tilde{\beta}'X_i)^2} X_i$$

$$+ \tilde{\alpha} \, \text{Plim} \frac{1}{n} \Sigma_{-} \frac{1 - 2\tilde{\alpha}\tilde{\gamma}_2(y_i - \tilde{\beta}'X_i)}{1 - \tilde{\alpha}(y_i - \tilde{\beta}'X_i) + \tilde{\alpha}^2\tilde{\gamma}_2(y_i - \tilde{\beta}'X_i)^2} X_i = 0$$

(Here, as before, $\tilde{a} = \text{Plim } a$, and similarly for the other parameters.)

In order to show the consistency of $\hat{\beta}$, we show that (4.5)-(4.7) are satisfied for $\tilde{\beta} = \beta$. Setting $\tilde{\beta} = \beta$, we obtain;

$$(4.8) \qquad \frac{1}{\tilde{\alpha}} - E|u| + E \left\{ \frac{|u| + 2\widetilde{\alpha\gamma}_2 u^2}{1 + \tilde{\alpha}|u| + \tilde{\alpha}^2 \tilde{\gamma}_2^2 u^2} \right\} = 0$$

$$(4.9) \qquad \frac{-1}{1 + \tilde{\gamma}_2} + \tilde{\alpha}^2 \, E \left\{ \frac{u^2}{1 + \tilde{\alpha}|u| + \tilde{\alpha}^2 \tilde{\gamma}_2^2 u^2} \right\} = 0$$

$$(4.10) \qquad \tilde{\alpha} P(u{>}0)\tilde{X} - \tilde{\alpha} P(u{<}0)\tilde{X} - \tilde{\alpha} P(u{>}0)E \left\{ \frac{1 + 2\widetilde{\alpha\gamma}_2 u}{1 + \tilde{\alpha}u + \tilde{\alpha}^2 \tilde{\gamma}_2 u^2} \Big| u{>}0 \right\} \tilde{X}$$

$$+ \tilde{\alpha} P(u{<}0) \, E \left\{ \frac{1 - 2\widetilde{\alpha\gamma}_2 u}{1 - \tilde{\alpha}u + \tilde{\alpha}^2 \tilde{\gamma}_2 u^2} \Big| u{<}0 \right\} \tilde{X} = 0$$

Equation (4.10) holds, because

$$P(\, u \le 0 \,) = P(\, u \ge 0 \,),$$

and because of the equality of 2 expectations:

$$(4.11) \qquad \text{Plim} \, \frac{1}{n_+} \sum_+ \frac{1 + 2\widetilde{\alpha\gamma}_2 u}{1 + \tilde{\alpha}u + \tilde{\alpha}^2 \tilde{\gamma}_2 u^2} X_1 = E \left\{ \frac{1 + 2\widetilde{\alpha\gamma}_2 u}{1 + \tilde{\alpha}u + \tilde{\alpha}^2 \tilde{\gamma}_2 u^2} X \Big| u{>}0 \right\}$$

$$= E \left\{ \frac{1 + 2\widetilde{\alpha\gamma}_2 u}{1 + \tilde{\alpha}u + \tilde{\alpha}^2 \tilde{\gamma}_2 u^2} \Big| u{>}0 \right\} \tilde{X} = E \left\{ \frac{1 - 2\widetilde{\alpha\gamma}_2 u}{1 - \tilde{\alpha}u + \tilde{\alpha}^2 \tilde{\gamma}_2 u^2} \Big| u{<}0 \right\} \tilde{X}$$

Therefore, no asymptotic bias results from the assumption of second order Sargan errors. Equations (4.8) and (4.9) will determine

the values for $\tilde{a}$, and $\tilde{\gamma}$. As long as they are positive, their actual values are not important.

4.3 The Tobit Model

Now we turn to the censored dependent variable model, and assume that there is only one regressor , a constant term. That is, we estimate the mean of the population. The model is defined as in (3.1), in which the u's are i.i.d with zero mean. We can only observe y's which are defined by $y_i = \max(0, y_i^*)$. We do not observe the y^*'s. The u's are in fact distributed as normal but we mistakenly assume them to be distributed as second-order Sargan. Therefore, the log likelihood function would be:

$$(4.12) \quad \ln L = m \ln \alpha - m \ln 4 - m \ln (1+\gamma_2) - \alpha \sum_{i=1}^{m} |y_i - \mu|$$

$$+ \sum_{i=1}^{m} \ln [1 + \alpha|y_i - \mu| + \alpha^2 \gamma_2 (y_i - \mu)^2]$$

$$+ (n-m) \ln F(-\mu)$$

where F is the second-order Sargan c.d.f.:

$$(4.13) \quad F(x) = \begin{cases} 1 - \dfrac{e^{-\alpha x}}{4(1+\gamma_2)} [2 + \alpha x + \gamma_2 + \gamma_2(1+\alpha x)^2] & x > 0 \\[4mm] \dfrac{e^{\alpha x}}{4(1+\gamma_2)} [2 - \alpha x + \gamma_2 + \gamma_2(1-\alpha x)^2] & x < 0 \end{cases}$$

Differentiating (4.12) with respect to all three parameters, and setting them equal to zero, results in:

$$(4.14) \quad \frac{\partial \ell nL}{\partial \hat{\alpha}} = \frac{m}{\hat{\alpha}} - \sum_{i=1}^{m} |y_i - \hat{\mu}| + \sum_{i=1}^{m} \frac{|y_i - \hat{\mu}| + 2\hat{\alpha}\hat{\gamma}_2(y_i - \hat{\mu})^2}{1 + \hat{\alpha}|y_i - \hat{\mu}| + \hat{\alpha}^2\hat{\gamma}_2(y_i - \hat{\mu})^2}$$

$$+ (n-m) \, W \, (\hat{\alpha}, \hat{\gamma}_2, \hat{\mu}) = 0$$

$$\frac{\partial \ell nL}{\partial \hat{\gamma}_2} = \frac{-m}{1 + \hat{\gamma}_2} + \sum_{i=1}^{m} \frac{\hat{\alpha}^2(y_i - \hat{\mu})^2}{1 + \hat{\alpha}|y_i - \hat{\mu}| + \hat{\alpha}^2\hat{\gamma}_2(y_i - \hat{\mu})^2} + (n-m) \, S \, (\hat{\alpha}, \hat{\gamma}_2, \hat{\mu}) = 0$$

(4.15)

$$\frac{\partial \ell nL}{\partial \hat{\mu}} = \hat{\alpha}(m_+ - m_-) - \hat{\alpha} \sum \frac{1 + 2\hat{\alpha}\hat{\gamma}_2(y_i - \hat{\mu})}{1 + \hat{\alpha}(y_i - \hat{\mu}) + \hat{\alpha}^2\hat{\gamma}_2(y_i - \hat{\mu})^2}$$

(4.16)

$$+ \hat{\alpha} \sum \frac{1 - 2\hat{\alpha}\hat{\gamma}_2(y_i - \hat{\mu})}{1 - \hat{\alpha}(y_i - \hat{\mu}) + \hat{\alpha}^2\hat{\gamma}_2(y_i - \hat{\mu})^2} + (n-m) \, T \, (\hat{\alpha}, \hat{\gamma}_2, \hat{\mu}) = 0$$

Now we divide (4.14)-(4.16) by m and take Plims to get;

$$(4.17) \quad \frac{1}{\tilde{\alpha}} - E \, \{|y - \tilde{\mu}| \mid y > 0\} + E \, \{\frac{|y - \tilde{\mu}| + 2\tilde{\alpha}\tilde{\gamma}_2(y - \tilde{\mu})^2}{1 + \tilde{\alpha}|y - \tilde{\mu}| + \tilde{\alpha}^2\tilde{\gamma}_2(y - \tilde{\mu})^2} \mid y > 0\}$$

$$+ \frac{P(y<0)}{P(y>0)} \, W \, (\tilde{\alpha}, \tilde{\gamma}_2, \tilde{\mu}) = 0$$

$$(4.18) \quad \frac{-1}{1 + \tilde{\gamma}_2} + E \, \{\frac{\tilde{\alpha}^2(y - \tilde{\mu})^2}{1 + \tilde{\alpha}|y - \tilde{\mu}| + \tilde{\alpha}^2\tilde{\gamma}_2(y - \tilde{\mu})^2} \mid y > 0\} + \frac{P(y<0)}{P(y>0)} \, S(\tilde{\alpha}, \tilde{\gamma}_2, \tilde{\mu}) = 0$$

(4.19) $\tilde{\alpha}P(y>\tilde{\mu}|y>0) - \tilde{\alpha}P(y<\tilde{\mu}|y>0)$

$$- \tilde{\alpha}P(y>\tilde{\mu}|y>0) \; E \; \{\frac{1+2\tilde{\alpha}\tilde{\gamma}_2(y-\tilde{\mu})}{1+\tilde{\alpha}(y-\tilde{\mu})+\tilde{\alpha}^2\tilde{\gamma}_2(y-\tilde{\mu})^2} \mid y>0, \; y>\tilde{\mu}\}$$

$$+ \tilde{\alpha}P(y<\tilde{\mu}|y>0) \; E \; \{\frac{1-2\tilde{\alpha}\tilde{\gamma}_2(y-\tilde{\mu})}{1-\tilde{\alpha}(y-\tilde{\mu})+\tilde{\alpha}^2\tilde{\gamma}_2(y-\tilde{\mu})^2} \mid y>0, \; y<\tilde{\mu}\}$$

$$+ \frac{P(y<0)}{P(y>0)} \; T \; (\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu}) \; = \; 0$$

Note that

$$Plim \; \frac{m+}{m} \; = \; P(y>\tilde{\mu} \mid y>0)$$

$$Plim \; \frac{m-}{m} \; = \; P(y<\tilde{\mu} \mid y>0)$$

$$Plim \; \frac{n-m}{m} \; = \; \frac{P(y<0)}{P(y>0)}$$

$$W(\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu}) \; = \; \frac{\partial F(-\tilde{\mu})/\partial\tilde{\alpha}}{F(-\tilde{\mu})} \; = \begin{cases} \dfrac{-\tilde{\mu}(1+\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2\tilde{\alpha}^2\tilde{\mu}^2)}{2+\tilde{\gamma}_2+\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2(1+\tilde{\alpha}\tilde{\mu})^2} & \tilde{\mu}>0 \\[4mm] \dfrac{-\tilde{\mu}e^{\tilde{\alpha}\tilde{\mu}}(1-\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2\tilde{\alpha}^2\tilde{\mu}^2)}{4(1+\tilde{\gamma}_2)-e^{\tilde{\alpha}\tilde{\mu}}[2+\tilde{\gamma}_2-\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2(1-\tilde{\alpha}\tilde{\mu})^2]} & \tilde{\mu}<0 \end{cases}$$

(4.20)

$$T(\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu}) = \frac{\partial F(-\tilde{\mu})/\tilde{\sigma}\mu}{F(-\tilde{\mu})} = \begin{cases} \dfrac{-\tilde{\alpha}(1+\widetilde{\alpha\mu}+\tilde{\alpha}^2\tilde{\gamma}_2\tilde{\mu}^2)}{2+\tilde{\gamma}_2+\widetilde{\alpha\mu}+\tilde{\gamma}_2(1+\widetilde{\alpha\mu})^2} & \tilde{\mu}>0 \\[4mm] \dfrac{-\tilde{\alpha}e^{\widetilde{\alpha\mu}}(1-\widetilde{\alpha\mu}+\tilde{\gamma}_2\tilde{\alpha}^2\tilde{\mu}^2)}{4(1+\tilde{\gamma}_2)-e^{\widetilde{\alpha\mu}}[2+\tilde{\gamma}_2-\widetilde{\alpha\mu}+\tilde{\gamma}_2(1-\widetilde{\alpha\mu})^2]} & \tilde{\mu}<0 \end{cases}$$

(4.21)

$$S(\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu}) = \frac{\partial F(-\tilde{\mu})/\partial\tilde{\gamma}_2}{F(-\tilde{\mu})} = \begin{cases} \dfrac{\widetilde{\alpha\mu}(1+\widetilde{\alpha\mu})}{(1+\tilde{\gamma}_2)[2+\tilde{\gamma}_2+\widetilde{\alpha\mu}+\tilde{\gamma}_2(1+\widetilde{\alpha\mu})^2]} & \tilde{\mu}>0 \\[4mm] \dfrac{\widetilde{\alpha\mu}e^{\widetilde{\alpha\mu}}(1-\widetilde{\alpha\mu})/(1+\tilde{\gamma}_2)}{4(1+\tilde{\gamma}_2)-e^{\widetilde{\alpha\mu}}[2+\tilde{\gamma}_2-\widetilde{\alpha\mu}+\tilde{\gamma}_2(1-\widetilde{\alpha\mu})^2]} & \tilde{\mu}<0 \end{cases}$$

(4.22)

It is easy to see that evaluating the probabilities is not hard, nor is calculating $E\{\ |\ y-\tilde{\mu}\ |\ \ |\ y\geq 0\ \}$ analytically, but the other expected values are not easy to calculate. Thus we have simulated 10,000 to 130,000 observations (depending upon the degree of censoring) from a normal distribution with mean zero and variance one, i.e., equations (4.17)-(4.19) are solved numerically. The results are shown in Table 4.1, under the heading "censored". There are two sets of results for the case of "variance known", with different values for a and γ_2.

The column labelled <u>censored (1)</u> is based on the values of a and γ_2 which correspond to variance equal to 1 and the fourth cumulant equal to zero, namely $a = 3.07638$, and $\gamma_2 = 2.15470$, and the column labelled <u>censored (2)</u> is based on values of a and γ_2 which correspond to

variance equal to 1 and the value of the density function (second order Sargan) equal to .3989 at X= 0, namely $a = 2.69500, \gamma_2 = 0.68884$. Missiakoulis(1983) in his study used $\gamma_2 = 1/3$ and $a = 2.12132$ which as we have shown in chapter 1 would not result in $\sigma^2 = 1$. He should have used $a = 2.44949$ instead.

In either case with the variance unknown, the biases are very small. In fact, except for large negative μ , there is almost no bias. In the case of " <u>unknown variance</u>", we let both a and γ_2 vary and be defined within the model. Here the bias is larger than in the known-variance case, but it is still not very large for samples that are at least half complete.

Also the results in both cases (σ known and σ unknown) are better than the results shown in Table 3.1, for the first order Sargan case. This makes the second order Sargan more attractive than the first order, although of course the bias should not be taken as the only means of selecting an alternative for normal c.d.f.

4.4 The Truncated Regression Model

Finally we discuss very briefly the truncated dependent variable case. Here we have the following, log of likelihood function;

$$(4.23) \quad \ln L = m \ln \alpha - m \ln 4 - m \ln (1+\gamma_2) - \alpha \sum_{i=1}^{m} | y_i - \mu |$$

$$+ \sum_{i=1}^{m} \ln [1+\alpha|y_i - \mu| + \alpha^2 \gamma_2 (y_i-\mu)^2] - m \ln F (\mu)$$

We take derivatives of (4.23) with respect to a , γ_2 , μ, set them equal

to zero, divide by m and then take probability limits:

$$(4.24) \quad \frac{1}{\tilde{\alpha}} - E\left\{|y-\mu|\,\Big|\,y>0\right\} + E\left\{\frac{|y-\tilde{\mu}|+2\tilde{\alpha}\tilde{\gamma}_2(y-\tilde{\mu})^2}{1+\tilde{\alpha}|y-\tilde{\mu}|+\tilde{\alpha}^2\tilde{\gamma}_2(y-\tilde{\mu})^2}\,\Big|\,y>0\right\}$$

$$- A(\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu}) = 0$$

$$(4.25) \quad \frac{-1}{1+\tilde{\gamma}_2} + \tilde{\alpha}^2\, E\left\{\frac{(y-\tilde{\mu})^2}{1+\tilde{\alpha}|y-\tilde{\mu}|+\tilde{\alpha}^2\tilde{\gamma}_2(y-\tilde{\mu})^2}\,\Big|\,y>0\right\}$$

$$- c(\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu}) = 0$$

$$(4.26) \quad \tilde{\alpha}P(y>\tilde{\mu}|y>0) - \tilde{\alpha}P(y<\tilde{\mu}|y>0)$$

$$-\tilde{\alpha}P(y>\tilde{\mu}|y>0)\, E\left\{\frac{1+2\tilde{\alpha}\tilde{\gamma}_2(y-\tilde{\mu})^2}{1+\tilde{\alpha}(y-\tilde{\mu})+\tilde{\alpha}^2\tilde{\gamma}_2(y-\tilde{\mu})^2}\,\Big|\,y>\tilde{\mu},y>0\right\}$$

$$+ \tilde{\alpha}P(y<\tilde{\mu}|y>0)\, E\left\{\frac{1-2\tilde{\alpha}\tilde{\gamma}_2(y-\tilde{\mu})^2}{1-\tilde{\alpha}(y-\tilde{\mu})+\tilde{\alpha}^2\tilde{\gamma}_2(y-\tilde{\mu})^2}\,\Big|\,y<\tilde{\mu},y>0\right\}$$

$$- B(\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu})$$

where

$$(4.27) \quad A(\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu}) = \begin{cases} \dfrac{\tilde{\mu}(1+\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2\tilde{\alpha}^2\tilde{\mu}^2)e^{-\tilde{\alpha}\tilde{\mu}}}{4(1+\tilde{\gamma}_2)-e^{-\tilde{\alpha}\tilde{\mu}}[2+\tilde{\gamma}_2+\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2(1+\tilde{\alpha}\tilde{\mu})^2]} & \tilde{\mu}>0 \\[20pt] \dfrac{\tilde{\mu}(1-\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2\tilde{\alpha}^2\tilde{\mu}^2)}{2+\tilde{\gamma}_2-\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2(1-\tilde{\alpha}\tilde{\mu})^2} & \mu<0 \end{cases}$$

$$
(4.28) \quad B(\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu}) = \begin{cases} \dfrac{\tilde{\alpha}e^{-\tilde{\alpha}\tilde{\mu}}(1+\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2\tilde{\alpha}^2\tilde{\mu}^2)}{4(1+\tilde{\gamma}_2)-e^{-\tilde{\alpha}\tilde{\mu}}[2+\tilde{\gamma}_2+\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2(1+\tilde{\alpha}\tilde{\mu})^2]} & \tilde{\mu}>0 \\[4mm] \dfrac{\tilde{\alpha}(1-\tilde{\alpha}\tilde{\mu}+\tilde{\alpha}^2\tilde{\gamma}_2\tilde{\mu}^2)}{2+\tilde{\gamma}_2-\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2(1-\tilde{\alpha}\tilde{\mu})^2} & \tilde{\mu}<0 \end{cases}
$$

$$
(4.29) \quad C(\tilde{\alpha},\tilde{\gamma}_2,\tilde{\mu}) = \begin{cases} \dfrac{-\tilde{\alpha}\tilde{\mu}e^{-\tilde{\alpha}\tilde{\mu}}(1+\tilde{\alpha}\tilde{\mu})/(1+\tilde{\gamma}_2)}{4(1+\tilde{\gamma}_2)-e^{-\tilde{\alpha}\tilde{\mu}}[2+\tilde{\gamma}_2+\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2(1+\tilde{\alpha}\tilde{\mu})^2]} & \tilde{\mu}>0 \\[4mm] \dfrac{\tilde{\alpha}\tilde{\mu}(\tilde{\alpha}\tilde{\mu}-1)/(1+\tilde{\gamma}_2)}{2+\tilde{\gamma}_2-\tilde{\alpha}\tilde{\mu}+\tilde{\gamma}_2(1-\tilde{\alpha}\tilde{\mu})^2} & \tilde{\mu}<0 \end{cases}
$$

As in the censored case we have calculated the necessary expected values by simulation, based on 10,000 to 130,000 replications. The results are shown in Table 4.1, and labelled <u>truncated</u> . Two sets of numbers are generated for the case of known variance.

As expected, the absolute biases are generally larger in the truncated regression model than in the censored regression model, and they are generally larger when the variance is unknown than when it is known. They are generally smaller than in the first-order Sargan case, sometimes substantially. So, at least from the standpoint of bias, it is safer to assume second-order Sargan than first-order Sargan.

Footnote

1) Incidently, we consider the second order Sargan distribution to have two free parameters (a and γ_2). This differs from Missiakoulis (1983) who constrains $\gamma_2 = 1/3$.

Table 4.1

Asymptotic Bias of Second Order Sargan 'MLE' when True Errors are N(0,1)

μ	Censored (1)	μ (Variance Known) Censored (2)	Truncated (1)	Truncated (2)	μ (Variance Unknown) Censored	Truncated
-2.8	-.07	-.29	*	*	.66	2.37
-2.4	.01	-.14	-8.56	*	.45	2.04
-2.0	.05	-.04	-0.91	-12.57	.27	1.72
-1.6	.05	.02	-.08	-2.17	.15	1.20
-1.2	.03	.04	.21	-.47	.03	.87
-.8	-.01	.04	.23	.01	-.07	.64
-.4	-.04	.02	.09	.09	-.12	.40
0.0	-.06	.02	-.09	.06	.01	.16
.4	.03	.01	-.11	.06	.01	.13
.8	.01	0.0	-.02	.05	0.0	0.7
1.2	0.0	0.0	.03	.02	0.0	.02
1.6	.01	0.0	.03	.01	0.0	.01
2.0	0.0	0.0	.01	0.0	0.0	-.01
2.4	0.0	0.0	0.0	-0.1	0.0	-0.1
2.8	0.0	0.0	-.01	-.01	0.0	-.01

Chapter 5

Bivariate Sargan Distribution

5.1 Introduction

So far we have considered only univariate Sargan distributions. In this chapter we extend this study to the bivariate case, in a way which could be generalized to the general n-variate case. If we view Sargan distributions as easily computable approximations to the normal, the logic for considering multivariate Sargan distributions is in fact much stronger than for univariate Sargan distributions, since the multivariate normal c.d.f is so much harder to calculate than the univariate normal c.d.f.

It is not immediately clear what a bivariate Sargan distribution is. Therefore, I begin by exploring different ways of defining such a bivariate density function, and some of their advantages as well as their shortcomings. Then we will approximate the bivariate density in a specific form. This is done on the basis of Stone's theorem (1962, p. 74), which shows that any function f(x) with certain characteristics can be uniformly approximated by functions of the form $e^{-a \, x} P(x)$, $P(x)$ being a polynomial.

We will also make some comparisons between the bivariate Sargan densities and their normal counterparts. Finally, we will look at the robustness of the estimators assuming such a bivariate density

for different models.

The plan of this chapter is as follows. In section 5.2 we discuss alternative possible definitions of a bivariate Sargan distribution, and choose one such definition. In section 5.3 we compare the bivariate Sargan density to the bivariate normal density (for various levels of correlation) to see how good an approximation we have. In section 5.4 we provide some evidence on the statistical relevance of the use of such an approximation to the hypothesized (true) error distribution function. For example, as the univariate case, the mistakenly assumed first-order Sargan distribution does not cause any asymptotic bias in the (seemingly unrelated) regression model, so long as the true distribution of the errors is a symmetric one. (A similarly strong result has already been shown for the univariate case.)

Finally, in section 5.5 we conclude that the bivariate Sargan distribution, as we have defined it, is an adequate and reasonable approximation for the bivariate normal distribution, at least when the correlation coefficient is not too large.

5.2 Definitions

There are several possible ways that one can proceed in order to construct a bivariate density (in our case a bivariate Sargan density.) We will investigate three ways of doing so, although not in great detail, and discuss their advantages as well as their problems.

i) Linear Transformation approach

Let us assume that X_1, X_2 are distributed independently according to the (first-order) Sargan with mean zero and variance one. Therefore, their joint density function will be;

$$(5.1) \qquad f(x_1, x_2) = \frac{16}{\alpha_1 \alpha_2} e^{-\alpha_1 |x_1| - \alpha_2 |x_2|} [1 + \alpha_1 |x_1|][1 + \alpha_2 |x_2|]$$

Now we define two other variables Y_1, Y_2 as linear functions of X's;

$$(5.2) \qquad \begin{aligned} Y_1 &= a_1 X_1 + a_2 X_2 \\ Y_2 &= b_1 X_1 + b_2 X_2 \end{aligned}$$

or simply,

$$(5.3) \qquad \begin{aligned} Y &= A X \\ A &= \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \end{aligned}$$

The joint density function of Y then becomes;

$$g(y_1, y_2) = \text{Jacobian} \cdot f(A^{-1} y) = ||A||^{-1} f(A^{-1} y)$$

$$(5.4) \qquad g(y_1, y_2) = ||A||^{-1} \frac{16}{\alpha_1 \alpha_2} e^{\frac{-\alpha_1}{\Delta}|b_2 y_1 - a_2 y_2| \frac{-\alpha_2}{\Delta}|a_1 y_2 - b_1 y_1|}$$

$$\cdot \left[1 + \frac{\alpha_1}{\Delta}|b_2 y_1 - a_2 y_2|\right]\left[1 + \frac{\alpha_2}{\Delta}|a_1 y_2 - b_1 y_1|\right]$$

with

(5.5)   COV $(Y_1, Y_2) = \Sigma = AA'$

$$\Delta = a_1b_2 - a_2b_1$$

This construction follows by analogy to the relationship between the bivariate and univariate normal, since the bivariate (or multivariate) normal can be defined as arising from linear transformation of independent univariate normals. We can accommodate any covariance matrix Σ by appropriate choice of A, and we could also allow for an arbitrary (non-zero) mean vector for Y. However, we dismiss this as a reasonable definition of bivariate Sargan because its marginals are not univariate Sargan.

ii) "Translation" approach

This is a general way of constructing a joint density with specified marginals, in this case Sargan marginals.

We start by assuming that X_1, X_2 have a bivariate normal density with correlation coefficient ρ. Let ϕ represent this density function. Now define:

$$Y_1 = F^{-1}\Phi(X_1)$$

(5.6)

$$Y_2 = F^{-1}\Phi(X_2)$$

where F is a univariate Sargan c.d.f., and Φ is the standard normal c.d.f. Then the marginal distributions of Y_1 and Y_2 are Sargan, by

construction. However, the joint density of Y_1 and Y_2 is messy:

$$(5.7) \qquad f(y_1, y_2) = \text{Jacobian } \phi[\Phi^{-1}F(y_1), \Phi^{-1}F(y_2)]$$

Thus there is no conceivable computational advantage of using such a bivariate Sargan density.

iii) Approximation Methods

Here we define the bivariate Sargan density as an approximation. To do this we will make use of a theorem due to Stone (1962, p. 74), that "<u>any continuous real function f, which is defined on the interval</u> $0 \leq x < \infty$ <u>and vanishes at infinity in the sense that</u> $\lim_{x \to \infty} f(x) = 0$, <u>can be uniformly approximated by functions of the</u> $e^{-a\,x}$ $P(x)$, <u>where P(x) is a polynomial</u>". Thus we start by assuming the density $f(x_1, x_2)$ to be a continuous function defined on the intervals $x_i \in [0, \infty)$, such that;

$$\lim_{x_1 \to \infty} f(x_1, x_2) = 0 \qquad\qquad \forall x_2$$

(5.8)

$$\lim_{x_2 \to \infty} f(x_1, x_2) = 0 \qquad\qquad \forall x_1$$

Following the argument made by Stone (1962, p. 75) we define two new variables ξ_1, ξ_2, as follows;

$$\xi_1 = e^{-\alpha_1 x_1}$$

(5.9)

$$\xi_2 = e^{-\alpha_2 x_2}$$

for arbitrary values of $x_1 > 0$, $x_2 > 0$

Therefore,

$$x_i = - \ell n \, \xi_i / a_i \qquad , i=1,2.$$

Notice that $\xi_i \in (0,1]$, and ξ_i goes to zero as x_i goes to infinity.

Now we define the function ϕ on $[0,1] \times [0,1]$, such that

$$\phi (\xi_1, \xi_2) = f(x_1, x_2) = f(- \ell n \, \xi_1 / a_1, - \ell n \, \xi_2 / a_2)$$

$$\phi (0, \xi_2) = 0 \qquad \forall \, \xi_2 \in (0,1]$$

(5.10)

$$\phi (\xi_1, 0) = 0 \qquad \forall \, \xi_1 \in (0,1]$$

$$\phi (0, 0) = 0$$

Then ϕ is a continuous function over $[0,1] \times [0,1]$, and therefore, it can

be uniformly approximated by a polynomial in ξ_1, ξ_2 (p. 69 of Stone),

as:

(5.11)
$$\begin{aligned}
&C_{00} + C_{10} \xi_1 + C_{01} \xi_2 \\
&+ C_{20} \xi_1^2 + C_{11} \xi_1 \xi_2 + C_{02} \xi_2^2 \\
&+ \dots \\
&+ C_{N0} \xi_1^N + C_{N-1,1} \xi_1^{N-1} \xi_2 + \dots + C_{0N} \xi_2^N
\end{aligned}$$

or simply

$$\sum_{0 < \beta_1 + \beta_2 < N} C_{\beta_1 \beta_2} \xi_1^{\beta_1} \xi_2^{\beta_2}$$

substituting for ξ_1, ξ_2 from (5.9) we get;

$$(5.12) \qquad \sum_{0 < \beta_1 + \beta_2 < N} C_{\beta_1 \beta_2} \ e^{-\alpha_1 \beta_1 x_1 - \alpha_2 \beta_2 x_2}$$

Now all we need to do is to approximate $\exp(-\alpha_i \beta_i x_i)$ uniformly, which can be done using lemma 1 of Stone (1962). The lemma says that $\exp(-\alpha_1 \beta_1 x_1)$ can be approximated uniformly by $\exp(-\alpha_1 x_1) \cdot P(x_1)$, where $P(x_1)$ is a polynomial in x_1 (which depends on x_1 and β_1), and similarily for $\exp(-\alpha_2 \beta_2 x_2)$. Multiplying and adding all these polynomials together we get another polynomial in x_1, x_2, that is;

$$(5.13) \qquad e^{-\alpha_1 x_1 - \alpha_2 x_2} Q(x_1, x_2)$$

where Q is a polynomial.

So far we have restricted ourselves to the non-negative values of X_i, $i=1,2$, but our study requires an approximation such that $x_i \in (-\infty, \infty)$. The question is whether this can be done. If we define for example,

$$\xi = \begin{cases} e^{-\alpha X} & X \geq 0 \\ \\ e^{\alpha X} & X < 0 \end{cases} = e^{-\alpha |X|}$$

another problem arises, since X will not have a one-to-one correspondance to ξ, or in other words X <u>is</u> <u>not</u> <u>a</u> <u>function</u> <u>of</u> ξ , since for any given value of ξ we can find two different values of X (which are the same in absolute value), that is:

$$\phi (\xi)= f(- \ln \xi / a) \text{ or } =f(\ln \xi / a)$$

and we do not know which to use.

In order to get around this problem we impose the symmetry condition on the density function, so,

$$f(- \ln \xi / a) =f(\ln \xi / a)$$

and $\phi (\xi)$ is well defined. Therefore, if f (X_1, X_2) is approximated uniformly by

$$e^{- a_1 X_1 - a_2 X_2} \bullet \text{(polynomial in } X_1, X_2)$$

over the range $[0, \infty)\times[0,\infty)$, by symmetry the negative part (third quadrant) can be uniformly approximated by $\exp(a_1 X_1 + a_2 X_2)\bullet$(polynomial in X_1, X_2) over $(- \infty ,0]\times(- \infty ,0]$. However, this still does not completely resolve the problem, since the other two quadrants $(-\infty ,0]\times[0,\infty)$ or $[0, \infty)\times(-\infty ,0]$, are still out of the picture.

From the standpoint of Stone's approximation theory, there seems to be no simple way to proceed without also making the second and fourth quadrants mirror image of the first. Suppose we do so, by assuming that

$$f(x_1, x_2) = f(-x_1, x_2) = f(x_1, -x_2) \; \forall \; x_1, x_2.$$

(Note that $f(x_1, x_2) = f(-x_1, -x_2)$ has already been imposed above.) Then $f(x_1, x_2)$ can be written as $f(|x_1|, |x_2|)$, and it follows directly that it can be approximated uniformly by

$$e^{-\alpha_1 |x_1| - \alpha_2 |x_2|} \cdot (\text{polynomial in } |x_1| \text{ and } |x_2|)$$

The considerable problem with this development is that if $f(x_1, x_2)$ can be written in terms of $|x_1|$ and $|x_2|$ only, as in $f(|x_1|, |x_2|)$ above, then X_1 and X_2 are necessarily uncorrelated. Thus an approximating density in terms of $|x_1|$, $|x_2|$, such that

$$(5.14) \quad f(x_1, x_2) = C \cdot e^{-\alpha_1 |x_1| - \alpha_2 |x_2|} (K_0 + K_1 |x_1| + K_2 |x_2| + K_3 |x_1| \cdot |x_2|)$$

where,

$$C = \frac{\alpha_1^2 \alpha_2^2}{4(\alpha_1 \alpha_2 K_0 + \alpha_2 K_1 + \alpha_1 K_2 + K_3)}$$

can not be alleged to be a good approximation to a bivariate density which exhibits non-zero correlation. Indeed, this is intuitively clear since the density (5.14) itself has zero correlation between X_1, and X_2. Other than that, it would be a reasonable choice from our perspective since it has other desirable properties.

Its marginal densities are as follows.

$$(5.15) \quad g(x_1) = \frac{C}{\alpha_2^2} e^{-\alpha_1 |x_1|} [\alpha_2 K_0 + K_2 + (\alpha_2 K_1 + K_3)|x_1|]$$

$$= k_1 e^{-\alpha_1 |x_1|} (1 + \alpha_1 |x_1|)$$

Note that if we let

$$\alpha_1 = \frac{\alpha_2 K_1 + K_3}{\alpha_2 K_0 + K_2} \quad , \text{ then}$$

$$k_1 = \frac{C}{\alpha_2^2}(\alpha_2 K_0 + K_2) = \frac{\alpha_1^2 \alpha_2^2 (\alpha_2 K_0 + K_2)}{8(\alpha_1 \alpha_2 K_0 + \alpha_1 K_2)\alpha_2^2} = \frac{\alpha_1}{8}$$

$$(5.16) \quad g(x_2) = \frac{C}{\alpha_1^2} e^{-\alpha_2 |x_2|} [\alpha_1 K_0 + K_1 + (\alpha_1 K_2 + K_3)|x_2|]$$

$$= k_2 e^{-\alpha_2 |x_2|} (1 + \alpha_2 |x_2|)$$

and again if we let

$$\alpha_2 = \frac{\alpha_1 K_2 + K_3}{\alpha_1 K_0 + K_1}$$

then

$$k_2 = \frac{C(\alpha_1 K_0 + K_1)}{\alpha_1^2} = \frac{\alpha_2}{8}$$

Thus the marginal densities are first-order Sargan, as we wish them to be. The means and variances are:

$$E(X_1) = 0$$

$$E(X_2) = 0$$

$$E(X_1^2) = k_1 \left\{ \int_{-\infty}^{0} e^{\alpha_1 x_1}(x_1^2 - \alpha_1 x_1^3)dx_1 \right.$$

$$\left. + \int_0^{\infty} e^{-\alpha_1 x_1}(x_1^2 + \alpha_1 x_1^3)dx_1 \right\}$$

$$= k_1 \left[\frac{2}{\alpha_1^3} + \frac{3!}{\alpha_1^3} + \frac{2}{\alpha_1^3} + \frac{3!}{\alpha_1^3} \right] = \frac{2}{\alpha_1^2}$$

$$E(X_2^2) = \frac{2}{\alpha_2^2}$$

(5.18) $$E(X_1, X_2) = COV(X_1, X_2) = 0$$

as noted above, and as shown directly in appendix A.

One way to allow dependence while maintaining Sargan marginals, though at the expense of the link with Stone's approximation theory, is to remove the absolute value from the cross-product in (5.14). Then we have the bivariate (Sargan) density,

(5.19) $$f(x_1, x_2) = A \cdot e^{-\alpha_1|x_1| - \alpha_2|x_2|}(K_0 + K_1|x_1| + K_2|x_2| + K_3 x_1 x_2)$$

where, $\iint f(X_1, X_2) = 1$ implies that;

(5.20) $$A = \frac{\alpha_1^2 \alpha_2^2}{4(\alpha_1 \alpha_2 K_0 + \alpha_2 K_1 + \alpha_1 K_2)}$$

This still results in marginal Sargan densities, if some restrictions are imposed. Explicitly,

$$(5.21) \quad g(x_1) = \frac{2A}{\alpha_2^2} (\alpha_2 K_0 + K_2 + \alpha_2 K_1 |x_1|) \, e^{-\alpha_1 |x_1|}$$

For this to be Sargan requires

$$(5.22) \quad \alpha_1 = \frac{\alpha_2 K_1}{\alpha_2 K_0 + K_2} \Rightarrow K_0 = \frac{K_1}{\alpha_1} - \frac{K_2}{\alpha_2}$$

Similarily, by symmetry of the argument

$$(5.23) \quad g(x_2) = \frac{2A}{\alpha_1^2} (\alpha_1 K_0 + K_1 + \alpha_1 K_2 |x_2|) \, e^{-\alpha_2 |x_2|}$$

Again in order for $g(X_2)$ to be a Sargan density requires;

$$(5.24) \quad \alpha_2 = \frac{\alpha_1 K_2}{\alpha_1 K_0 + K_1} \quad , \text{ which implies } \quad K_0 = \frac{K_2}{\alpha_2} - \frac{K_1}{\alpha_1}$$

(5.23) and (5.24) together imply that

$$K_0 = 0$$
(5.25)

$$K_1 / a_1 = K_2 / a_2$$

That is, to have Sargan marginal densities, the above equalities should hold. Imposing the second requirement as

$$(5.26) \qquad K_1 / a_1 = K_2 / a_2 = \theta \qquad \text{or } K_1 = \theta a_1 \qquad , K_2 = \theta a_2$$

θ being some parameter, then $A = a_1 a_2 / 8 \theta$. Substituting back the values of the parameters into the bivariate density function and its marginals we get very simple results

$$(5.27) \quad f(x_1, x_2) = \frac{\alpha_1 \alpha_2}{8\theta} e^{-\alpha_1 |x_1| - \alpha_2 |x_2|} [\theta(\alpha_1 |x_1| + \alpha_2 |x_2|) + K_3 x_1 x_2]$$

$$g(x_1) = \frac{\alpha_1}{4} e^{-\alpha_1 |x_1|} (1 + \alpha_1 |x_1|)$$

$$(5.28)$$

$$g(x_2) = \frac{\alpha_2}{4} e^{-\alpha_2 |x_2|} (1 + \alpha_2 |x_2|)$$

The principal advantage of (5.27) over (5.14) is that it does not imply zero covariance. In appendix A it is shown that $\text{cov}(X_1, X_2)$ depends on the values of the parameters, in particular K_3. Explicitly,

$$(5.29) \quad \text{COV } (X_1, X_2) = E(X_1, X_2) = \frac{2K_3}{\alpha_1^2 \alpha_2^2 \theta}$$

$$E(X_1) = 0 \qquad\qquad E(X_1^2) = 4 / \alpha_1^2$$

$$(5.30)$$

$$E(X_2) = 0 \qquad\qquad E(X_2^2) = 4 / \alpha_2^2$$

There is still one problem remaining to be resolved, namely that $f(x_1, x_2)$ in (5.27) need not to be non-negative for all values of

x_1, x_2 and the parameters. We will solve this problem by including another term which involves the cross-product of absolute values of x_1 and x_2, and therefore rewrite the joint density as follows:

$$f(x_1, x_2) = B \cdot e^{-\alpha_1 |x_1| - \alpha_2 |x_2|} \{K_0 + K_1 |x_1| + K_2 |x_2| + K_3 x_1 x_2 + K_4 |x_1||x_2|\}$$

(5.31)

with $K_3 < K_4$ to ensure $f(x_1, x_2) \geq 0$. Then B should be such that;

$$\iint f(x_1, x_2) \, dx_1 dx_2 = 1$$

Therefore,

$$(5.32) \quad B = \frac{\alpha_1^2 \alpha_2^2}{4(\alpha_1 \alpha_2 K_0 + \alpha_2 K_1 + \alpha_1 K_2 + K_4)}$$

Its marginal densities are exactly Sargan, given certain restrictions. They are:

$$g(x_1) = \frac{2Ae^{-\alpha_1 |x_1|}}{\alpha_2^2} \{\alpha_2 K_0 + K_2 + (\alpha_2 K_1 + K_4)|x_1|\}$$

$$g(x_2) = \frac{2Ae^{-\alpha_2 |x_2|}}{\alpha_1^2} \{\alpha_1 K_0 + K_1 + (\alpha_1 K_2 + K_4)|x_1|\}$$

For these to be Sargan require

$$\alpha_1 = \frac{\alpha_2 K_1 + K_4}{\alpha_2 K_0 + K_2} \implies K_0 = \frac{K_1}{\alpha_1} - \frac{K_2}{\alpha_2} + \frac{K_4}{\alpha_1 \alpha_2}$$

(5.33)

$$\alpha_2 = \frac{\alpha_1 K_2 + K_4}{\alpha_1 K_0 + K_1} \implies K_0 = \frac{K_2}{\alpha_2} - \frac{K_1}{\alpha_1} + \frac{K_4}{\alpha_1 \alpha_2}$$

Thus they require

$$\frac{K_2}{\alpha_2} = \frac{K_1}{\alpha_1} \; , \; = \theta \text{ (say)}, \quad K_0 = \frac{K_4}{\alpha_1 \alpha_2}$$

Then

$$\alpha_1 = \frac{K_1}{\theta} \; , \; \alpha_2 = \frac{K_2}{\alpha_2} \qquad \text{or}$$

(5.34) $\quad K_1 = \alpha_1 \theta \; , \; K_2 = \alpha_2 \theta \; , \; K_4 = \alpha_1 \alpha_2 K_0$

With these substitutions

$$(5.35) \quad B = \frac{\alpha_1^2 \alpha_2^2}{8\alpha_1 \alpha_2 (K_0 + \theta)} = \frac{\alpha_1 \alpha_2}{8(K_0 + \theta)}$$

and

$$(5.36) \quad f(x_1, x_2) = \frac{\alpha_1 \alpha_2}{8(K_0 + \theta)} e^{-\alpha_1 |x_1| - \alpha_2 |x_2|} [K_0 + \alpha_1 \theta |x_1|$$

$$+ \alpha_2 \theta |x_2| + K_3 x_1 x_2 + \alpha_1 \alpha_2 K_0 |x_1| |x_2|]$$

The marginal densities become univariate Sargan:

$$g_1(x_1) = \frac{\alpha_2}{4} e^{-\alpha_2|x_2|}(1+\alpha_2|x_2|)$$

(5.37)

$$g_2(x_2) = \frac{\alpha_2}{4} e^{-\alpha_2|x_2|}(1+\alpha_2|x_2|)$$

with means zero and variances being $4/\alpha_1^2$, $4/\alpha_2^2$ respectively for x_1 and x_2. Next we calculate the covariance

$$(5.38) \quad COV\ (X_1,X_2) = E(X_1,X_2) = \frac{2K_3}{\alpha_1^2\alpha_2^2(K_0+\theta)}$$

The correlation coefficient ρ therefore becomes;

$$(5.39) \quad \rho = \frac{COV\ (X_1,X_2)}{\sqrt{VAR(X_1)VAR(X_2)}} = \frac{K_3}{2\alpha_1\alpha_2(K_0+\theta)}$$

Recall that in (5.36) we need to have $K_3 \leq \alpha_1\alpha_2K_0$ to ensure that the density is non-negative for all X_1, X_2. This puts some restriction on the possible correlations that can be allowed. For example, if $\theta = 0$ then $\rho \leq 1/2$; if $\theta > 0$ not even $\rho = 1/2$ can be attained. This seems to be the main shortcoming of the bivariate Sargan distribution as defined in (5.36), and it casts some doubt on its ability to approximate well distributions with large positive correlation. In appendix B we also derive the moment generating

function, which implies the following moments.

$$\frac{[n!m!+(n+1)!(m+1)!]\ K_0 + [(n+1)!\ m! + n!\ (m+1)!]\ \theta}{2\alpha_1^{\,n}\alpha_2^{\,m}(K_0+\theta)}$$

if n,m both even

$$\mu_{nm} = \quad 0 \qquad\qquad \text{if one is odd and the other one is even}$$

$$\frac{(n+1)!\ (m+1)!\ K_3}{2\alpha_1^{\,n+1}\ \alpha_2^{\,m+1}\ (K_0+\theta)}$$

if n,m both are odd

(5.40)

Also the cumulative distribution function is equal to:

$$\frac{e^{\alpha_1 z_1}}{8C_4}\ \{\ -e^{-\alpha_2 z_2}[C_6-\alpha_1 C_7 z_1+\alpha_2 C_7 z_2-C_5 z_1 z_2] + 4C_4-2\alpha_1 C_4 z_1\}$$

$$z_2 < 0,\ z_1 > 0 \text{ or } z_1 < 0,\ z_2 > 0$$

$$F(z_1 z_2)$$

$$\frac{e^{\alpha_1 z_1 + \alpha_2 z_2}}{8C_4}\ \{C_1-\alpha_1 C_2 z_1-\alpha_2 C_2 z_2+C_3 z_1 z_2\} \qquad z_1,z_2 < 0$$

$$\frac{1}{8C_4}\ \{e^{-\alpha_1 z_1 - \alpha_2 z_2}[C_1+\alpha_1 C_2 z_1+\alpha_2 C_2 z_2+C_3 z_1 z_2] + C_1$$

(5.41)

$$-e^{-\alpha_1 z_1}[C_1+\alpha_1 C_2 z_1] - e^{-\alpha_2 z_2}[C_1+\alpha_2 C_2 z_2]\} \qquad z_1,z_2 > 0$$

where

$$C_1 = 2\alpha_1\alpha_2(K_0+\theta)+K_3$$

$$C_2 = \alpha_1 \alpha_2 (K_0 + \theta) + K_3$$

$$C_3 = \alpha_1 \alpha_2 (K_3 + \alpha_1 \alpha_2 K_0)$$

$$C_4 = \alpha_1 \alpha_2 (K_0 + \theta)$$

$$C_5 = \alpha_1 \alpha_2 (\alpha_1 \alpha_2 K_0 - K_3)$$

$$C_6 = 2\alpha_1 \alpha_2 (K_0 + \theta) - K_3$$

$$C_7 = \alpha_1 \alpha_2 (K_0 + \theta) - K_3$$

The bivariate Sargan density (5.36) can be easily extended to the multivariate case, but we will not pursue the details here. Also, it is possible to extend the (first-order) bivariate Sargan density (5.36) to the second-order or higher order cases. In the second-order case, we include all terms in the polynomial in (5.36) plus X_1^2 and X_2^2. We could also include terms like $X_1^2 X_2$, $X_1 X_2^2$ and $X_1^2 X_2^2$, but then we would not get second-order Sargan marginals, so we will not do so. Thus we define the second-order bivariate Sargan distribution to have a density of the form

$$(5.42) \quad f(x_1, x_2) = K \cdot e^{-\alpha_1 |x_1| - \alpha_2 |x_2|} [K_0 + K_1 |x_1| + K_2 |x_2| + K_3 x_1 x_2$$
$$+ K_4 |x_1| |x_2| + K_5 x_1^2 + K_6 x_2^2]$$

To ensure the non-negativity of f for all the values of X_1 and X_2 we have to have $K_3 \leq K_4$. The parameter K is defined by the condition

$\iint f(X_1, X_2) = 1$ which results in

$$(5.43) \quad K = \frac{\alpha_1^3 \alpha_2^3}{4(\alpha_1^2 \alpha_2^2 K_0 + \alpha_1 \alpha_2^2 K_1 + \alpha_1^2 \alpha_2 K_2 + \alpha_1 \alpha_2 K_4 + 2\alpha_2^2 K_5 + 2\alpha_1^2 K_6)}$$

Its marginal densities are, as we will see, exactly second-order Sargan given certain restrictions. They are derived in the appendix B.

Thus we have

$$(5.44) \quad g_1(x_1) = \frac{\alpha_1 e^{-\alpha_1 |x_1|}}{2(1+\gamma_1 + 2\gamma_2)} (1 + \alpha_1 \gamma_1 |x_1| + \alpha_1^2 \gamma_2 x_1^2)$$

$$(5.45) \quad g_2(x_2) = \frac{\alpha_2 e^{-\alpha_2 |x_2|}}{2(1+\xi_1 + 2\xi_2)} (1 + \alpha_2 \xi_1 |x_2| + \alpha_2^2 \xi_2 x_2^2)$$

where

$$\gamma_1 = \frac{\alpha_2^2 K_1 + \alpha_2 K_4}{\alpha_1 (\alpha_2^2 K_0 + \alpha_2 K_2 + 2L_6)} \quad , \quad \gamma_2 = \frac{\alpha_2^2 K_5}{\alpha_1^2 (\alpha_2^2 K_0 + \alpha_2 K_2 + 2K_6)}$$

$$\xi_1 = \frac{\alpha_1^2 K_2 + \alpha_1 K_4}{\alpha_2 (\alpha_1^2 K_0 + \alpha_1 K_1 + 2K_5)} \quad , \quad \xi_2 = \frac{\alpha_1^2 K_6}{\alpha_2^2 (\alpha_1^2 K_0 + \alpha_1 K_1 + 2K_5)}$$

and their variances:

$$\text{Var}(X_1) = \frac{2 + 6\gamma_1 + 24\gamma_2}{\alpha_1^2 (1 + \gamma_1 + 2\gamma_2)}$$

$$(5.46)$$

$$\text{Var}(X_2) = \frac{2 + 6\xi_1 + 24\xi_2}{\alpha_2^2 (1 + \xi_1 + 2\xi_2)}$$

Note $\xi_1 = \gamma_1 = 1$ is required to have continuous first and second partials.

Now according to the appendix C, the joint density will be;

$$(5.47) \quad f(x_1, x_2) = \frac{\alpha_1 \alpha_2}{8\delta} e^{-\alpha_1|x_1|-\alpha_2|x_2|} [K_0 + \alpha_1 v_1|x_1| + \alpha_2 v_2|x_2| + K_3 x_1 x_2$$

$$+ \alpha_1 \alpha_2 v_3 |x_1||x_2| + \alpha_1^2 \delta v_4 x_1^2 + \alpha_2^2 \delta v_5 x_2^2]$$

with

$$(5.48) \quad v_1 = \delta\left(\frac{1}{1+\xi_2} - 2\frac{\gamma_2}{1+\gamma_2}\right) - K_0 = \delta\left(\frac{1}{1+\xi_2} - 2v_4\right) - K_0$$

$$(5.49) \quad v_2 = \delta\left(\frac{1}{1+\gamma_2} - 2\frac{\xi_2}{1+\xi_2}\right) - K_0 = \delta\left(\frac{1}{1+\gamma_2} - 2v_5\right) - K_0$$

$$(5.50) \quad v_3 = \delta\left(\frac{\xi_2}{1+\xi_2} + \frac{\gamma_2}{1+\gamma_2}\right) + K_0 = \delta(v_5 + v_4) + K_0$$

$$(5.51) \quad v_4 = \frac{\gamma_2}{1+\gamma_2} \quad , \quad v_5 = \frac{\xi_2}{1+\xi_2}$$

$$(5.52) \quad \delta = K_5 \frac{(1+\gamma_2)}{\alpha_1^2 \gamma_2} = K_6 \frac{(1+\xi_2)}{\alpha_2^2 \xi_2} = \frac{K_5}{\alpha_1^2 v_4} = \frac{K_6}{\alpha_2^2 v_5}$$

$$(5.53) \quad \text{COV}(X_1, X_2) = E(X_1, X_2) = \frac{2K_3}{\delta \alpha_1^2 \alpha_2^2}$$

$$(5.54) \quad \rho = \frac{K_3}{2\alpha_1 \alpha_2 \delta} \sqrt{\frac{(1+\gamma_2)(1+\xi_2)}{(1+3\gamma_2)(1+3\xi_2)}}$$

The main problem with the second-order bivariate Sargan distribution is that it contains a large number of parameters. A bivariate normal with zero mean contains 3 parameters. The first-order bivariate Sargan (5.36) contains 4 parameters. The second-order bivariate Sargan contains 10 parameters (a_1, a_2, δ, K_0, K_3, v_1, v_2, v_3, v_4, v_5). This may be more than is reasonable.

The alternative second-order bivariate Sargan density that was considered is the following:

$$
\begin{aligned}
f(x_1,x_2) = A \cdot e^{-\alpha_1|x_1|-\alpha_2|x_2|} & [K_0 + K_1|x_1| + K_2|x_2| + K_3 x_1 x_2 + K_4|x_1||x_2| \\
& + K_5 x_1^2 + K_6 x_2^2 + K_7 x_1^2 x_2 + K_8 x_1^2|x_2| \\
& + K_9 x_1 x_2^2 + K_{10}|x_1|x_2^2 + K_{11} x_1^2 x_2^2]
\end{aligned}
$$

(5.55)

with the following assumptions;

$$K_3 \leq K_4$$

$$K_7 \leq K_8$$

$$K_9 \leq K_{10}.$$

Here again A has to be equal to:

(5.56) $A = a_1^3 a_2^3 / 4D$

where,

$$D = \alpha_2^2(\alpha_1^2 K_0 + \alpha_1 K_1 + 2K_5) + \alpha_2(\alpha_1^2 K_2 + \alpha_1 K_4 + 2K_8) + 2(\alpha_1^2 K_6 + \alpha_1 K_{10} + K_{11})$$

(5.57)

to make $f(x_1, x_2)$ a density function. The problem with this joint density function is that it does not generate the second-order Sargan marginals. Rather

$$(5.58) \quad g_1(x_1) = \begin{cases} \dfrac{2Ae^{-\alpha_1|x_1|}}{\alpha_2^3} [r_1 + (r_2 + 2(K_9 + K_{10}))x_1 + r_3 x_1^2] & x_1 > 0 \\[4ex] \dfrac{2Ae^{-\alpha_1|x_1|}}{\alpha_2^3} [r_1 - (r_2 + 2(K_9 - K_{10}))x_1 + r_3 x_1^2] & x_1 < 0 \end{cases}$$

$$(5.59) \quad g_2(x_2) = \begin{cases} \dfrac{2Ae^{-\alpha_2|x_2|}}{\alpha_1^3} \{r_4 + (r_5 + 2(K_7 + K_8))x_2 + r_6 x_2^2\} & x_2 > 0 \\[4ex] \dfrac{2Ae^{-\alpha_2|x_2|}}{\alpha_1^3} \{r_4 - (r_5 + 2(K_8 - K_7))x_2 + r_6 x_2^2\} & x_2 < 0 \end{cases}$$

where,

$$r_1 = \alpha_2^2 K_0 + \alpha_2 K_2 + 2K_6$$

$$r_2 = \alpha_2^2 K_1 + \alpha_2 K_4$$

$$r_3 = \alpha_2^2 K_5 + \alpha_2 K_8 + 2K_{11}$$

$$r_4 = \alpha_1^2 K_0 + \alpha_1 K_1 + 2K_5$$

$$r_5 = \alpha_1^2 K_2 + \alpha_1 K_4$$

$$r_6 = \alpha_1^2 K_6 + \alpha_1 K_{10} + 2K_{11}$$

Also, this density contains even more parameters than (5.47). For both reasons we will not consider it further.

5.3 Density Comparisons

In this section we will provide tables of first-order bivariate Sargan and bivariate normal densities. In order to be able to make meaningful comparisons we have imposed some restrictions on both densities. These restrictions are;

(i) means equal to zero, variances equal to one .

(ii) same correlation coefficient, ρ .

(iii) same densities at $x_1 = x_2 = 0$.

These restrictions are sufficient to determine uniquely the bivariate Sargan distribution which we will compare to the bivariate normal. The reason is that, although the Sargan density (5.36) appears to depend on five parameters (α_1, α_2, K_0, K_3, θ), it can in fact be written in terms of only four parameters. Let ρ be defined as in (5.39) and define

$$(5.60) \quad S_0 = f(0,0) = \frac{\alpha_1 \alpha_2 K_0}{8(K_0+\theta)}$$

then

$$f(x_1,x_2) = \frac{\alpha_1 \alpha_2}{8(K_0+\theta)} e^{-\alpha_1|x_1|-\alpha_2|x_2|} [K_0+\alpha_1\theta|x_1|+\alpha_2\theta|x_2|+K_3 x_1 x_2$$

$$+ \alpha_1 \alpha_2 K_0|x_1 x_2|]$$

$$= \frac{\alpha_1\alpha_2 e^{-\alpha_1|x_1|-\alpha_2|x_2|}}{8(K_0+\theta)} \{(K_0+\theta)[\alpha_1|x_1|+\alpha_2|x_2|]$$

$$+ K_0[1-\alpha_1|x_1|-\alpha_2|x_2|+\alpha_1\alpha_2|x_1 x_2|]$$

$$+ K_3 x_1 x_2\}$$

$$(5.61) \quad = e^{-\alpha_1|x_1|-\alpha_2|x_2|}\{\frac{\alpha_1\alpha_2}{8}[\alpha_1|x_1|+\alpha_2|x_2|]$$

$$+ S_0[1-\alpha_1|x_1|-\alpha_2|x_2|+\alpha_1\alpha_2|x_1 x_2|]$$

$$+ \frac{\alpha_1^2\alpha_2^2}{4}\rho x_1 x_2\}$$

The imposition of unitary variances implies $a_1 = a_2 = 2$, so that the Sargan distribution is uniquely determined once we pick a value of the correlation (ρ) and of the density at $X_1 = X_2 = 0$ (S_0).

We therefore provide in Tables 5.1-5.7 a comparison of the Sargan and normal densities. Each table corresponds to a different value of ρ. For $X_1 = \{-3.0,-2.5,-2.0,-1.5,-1.0,-0.5,0.0,0.5,1.0,1.5,2.0, 2.5,3.0\}$ and $X_2 = \{0.0,0.5,1.0,1.5,2.0,2.5,3.0\}$, we provide the normal (top entry) and Sargan (bottom entry) density. (Negative entries for X_2 are unnecessary because of symmetry; e.g., $f(x_1,-1.0) = f(-x_1,1.0)$ and $f(x_1, x_2) = f(x_2, x_1)$.)

The agreement between the Sargan and normal is rather close.

The largest differences are near the turning points of the distribution
($|X|$ between 1.0 and 2.0, say), but even these are not very large.
Thus, for the values of ρ considered, the Sargan distribution appears
to be a fairly good approximation to the normal. However, it should be
noted that the range of ρ which we consider is rather limited. As
indicated in the discussion following (5.39) above, we need to restrict
the correlation (ρ) to ensure that the Sargan density is non-negative
for all X_1 and X_2. We can not expect a Sargan density to approximate
well a bivariate normal with high correlation, at least not unless we
allow negative "density" in the tails of the approximation.

5.4 A Simple Seemingly Unrelated Regression Model

In the previous section we saw that the bivariate Sargan
distribution is a reasonably accurate approximation to the bivariate
normal. However, this fact does not provide direct evidence on the
relevant question of whether estimators based on the bivariate Sargan
distribution will be robust to normal errors. This question was
addressed in chapters 3 and 4 in the univariate case, for a variety of
models, and the answer was (not surprisingly) model dependent. It is
reasonable to expect that the same will be true in the bivariate case;
the asymptotic bias that results from assuming Sargan errors when the
errors are actually normal will depend on the model. Rather than
become involved in the extensive calculations which were done in the
univariate case, however, we will simply ask whether this bias is ever
zero. It was, in the univariate case, in the linear regression model.

Thus it is reasonable to conjecture that this bias may be zero, in the bivariate case, in the seemingly unrelated regression model.

For simplicity and tractability we restrict ourselves to only two equations and only one regressor (a constant term), in each equation. We show that the estimates based on the bivariate Sargan distribution are indeed consistent, regardless of the form of the true distribution of error terms, so long as it is symmetric around zero.

Thus we consider the simple model

$$y_1 = \mu_1 + \epsilon_1$$

(5.62)

$$y_2 = \mu_2 + \epsilon_2$$

where y_i is Tx1, ϵ_i has mean zero and covariance structure:

$$E(\epsilon_{ti}, \epsilon_{sj}) = \begin{cases} 0 & \text{if } t \neq s \\ \\ \sigma_{ij} & \text{if } t=s \end{cases} \quad i,j=1,2$$

Notice that we have not yet specified the distribution function of the errors. Here, as in chapters 3 and 4, we ask the question of what happens if we assume a bivariate Sargan distribution when in fact the true distribution is something else (such as bivariate normal). We form the (Sargan) log likelihood function;

$$(5.63) \quad \mathcal{L} = M \ln B + \sum_{m=1}^{M} \ln f(y_{m1} - \mu_1, \ y_{m2} - \mu_2)$$

or

$$\mathcal{L} = M\ell nB - \alpha_1 \sum_{m=1}^{M} |y_m - \mu_1| - \alpha_2 \sum_{m=1}^{M} |y_{m2} - \mu_2|$$

$$+ \sum_{m=1}^{M} \ell n \{K_0 + \alpha_1 \theta |y_{m1} - \mu_1| + \alpha_2 \theta |y_{m2} - \mu_2| + K_3(y_{m1} - \mu_1)(y_{m2} - \mu_2)\}$$

(5.64)
$$+ \alpha_1 \alpha_2 K_0 |(y_{m1} - \mu_1)(y_{m2} - \mu_2)| \} \ .$$

where B is defined as (5.35).

To express its derivatives with respect to all parameters, we define:

(5.65) $\quad D_1 = K_0 + \alpha_1 \theta(y_{m1} - \mu_1) + \alpha_2 \theta(y_{m2} - \mu_2) + (K_3 + \alpha_1 \alpha_2 K_0)(y_{m1} - \mu_1)(y_{m2} - \mu_2)$

(5.66) $\quad D_2 = K_0 + \alpha_1 \theta(y_{m1} - \mu_1) - \alpha_2 \theta(y_{m2} - \mu_2) + (K_3 - \alpha_1 \alpha_2 K_0)(y_{m1} - \mu_1)(y_{m2} - \mu_2)$

(5.67) $\quad D_3 = K_0 - \alpha_1 \theta(y_{m1} - \mu_1) + \alpha_2 \theta(y_{m2} - \mu_2) + (K_3 - \alpha_1 \alpha_2 K_0)(y_{m1} - \mu_1)(y_{m2} - \mu_2)$

(5.68) $\quad D_4 = K_0 - \alpha_1 \theta(y_{m1} - \mu_1) - \alpha_2 \theta(y_{m2} - \mu_2) + (K_3 + \alpha_1 \alpha_2 K_0)(y_{m1} - \mu_1)(y_{m2} - \mu_2)$

(5.69) $\quad D_5 = K_0 + \alpha_1 \theta |y_{m1} - \mu_1| + \alpha_2 \theta |y_{m2} - \mu_2| + K_3(y_{m1} - \mu_1)(y_{m2} - \mu_2)$

$$+ \alpha_1 \alpha_2 K_0 |(y_{m1} - \mu_1)(y_{m2} - \mu_2)|$$

Then , the first order conditions defining the maximum likelihood estimates are;

$$\frac{\partial \pounds}{\partial \mu_1} = \alpha_1(M_{1+} - M_{1-}) - \sum^{++} \frac{\alpha_1 \theta + (K_3 + \alpha_1 \alpha_2 K_0)(y_{m2} - \mu_2)}{D_1}$$

$$- \sum^{+-} \frac{\alpha_1 \theta + (K_3 - \alpha_1 \alpha_2 K_0)(y_{m2} - \mu_2)}{D_2}$$

$$+ \sum^{-+} \frac{a_1 \theta - (K_3 - \alpha_1 \alpha_2 K_0)(y_{m2} - \mu_2)}{D_3}$$

(5.70)
$$+ \sum^{++} \frac{\alpha_1 \theta - (K_3 + \alpha_1 \alpha_2 K_0)(y_{m2} - \mu_2)}{D_4} = 0$$

(5.71)
$$\frac{\partial \pounds}{\partial \mu_2} = \alpha_2(M_{2+} - M_{2-}) - \sum^{++} \frac{\alpha_2 \theta + (K_3 + \alpha_1 \alpha_2 K_0)(y_{m1} - \mu_1)}{D_1}$$

$$+ \sum^{+-} \frac{\alpha_2 \theta - (K_3 - \alpha_1 \alpha_2 K_0)(y_{m1} - \mu_1)}{D_2}$$

$$- \sum^{-+} \frac{\alpha_2 \theta + (K_3 - \alpha_1 \alpha_2 K_0)(y_{m1} - \mu_1)}{D_3}$$

$$+ \sum^{+-} \frac{\alpha_2 \theta - (K_3 + \alpha_1 \alpha_2 K_0)(y_{m1} - \mu_1)}{D_4} = 0$$

(5.72)
$$\frac{\partial \pounds}{\partial \alpha_1} = \frac{M}{\alpha_1} - \sum_{m=1}^{M} |y_{m1} - \mu_1| +$$

$$\sum_{m=1}^{M} \frac{\theta |y_{m1} - \mu_1| + \alpha_2 K_0 |(y_{m1} - \mu_1)(y_{m2} - \mu_2)|}{D_5} = 0$$

$$(5.73) \quad \frac{\partial \pounds}{\partial \alpha_2} = \frac{M}{\alpha_2} - \sum_{m=1}^{M} | y_{m2} - \mu_2 | +$$

$$\sum_{m=1}^{M} \frac{\theta | y_{m2} - \mu_2 | + \alpha_1 K_0 | (y_{m1} - \mu_1)(y_{m2} - \mu_2) |}{D_5} = 0$$

$$(5.74) \quad \frac{\partial \pounds}{\partial K_0} = \frac{-M}{K_0 + \theta} + \sum_{m=1}^{M} \frac{1 + \alpha_1 \alpha_2 | (y_{m1} - \mu_1)(y_{m2} - \mu_2) |}{D_5} = 0$$

$$(5.75) \quad \frac{\partial \pounds}{\partial \theta} = \frac{-M}{K_0 + \theta} + \sum_{m=1}^{M} \frac{\alpha_1 | y_{m1} - \mu_1 | + \alpha_2 | y_{m2} - \mu_2 |}{D_5} = 0$$

$$(5.76) \quad \frac{\partial \pounds}{\partial K_3} = \sum_{m=1}^{M} \frac{(y_{m1} - \mu_1)(y_{m2} - \mu_2)}{D_5} = 0$$

(where, $\underline{M}_{++}$ is the number of terms in $^{+}\Sigma^{+}$ and similarly for $\underline{M}_{+}$, $\underline{M}_{+-}$, $\underline{M}_{-}$.)

Dividing these derivatives by M, and taking probability limits we have (denoting the probability limits of α_i, μ_i, K_0, K_3, θ by $\tilde{\alpha}_i$, $\tilde{\mu}_i$, $\tilde{K}_0$, $\tilde{K}_3$, $\tilde{\theta}$ respectfully):

$$(5.77) \quad \tilde{\alpha}_1 \left\{ \text{Plim} \frac{M_{1+}}{M} - \text{Plim} \frac{M_{1-}}{M} \right\}$$

$$- \text{Plim} \frac{M_{++}}{M} \text{Plim} \frac{1}{M_{++}} \sum^{++} \frac{\tilde{\alpha}_1 \tilde{\theta} + (\tilde{K}_3 + \tilde{\alpha}_1 \tilde{\alpha}_2 \tilde{K}_0)(y_{m2} - \tilde{\mu}_2)}{\tilde{D}_1}$$

$$- \text{Plim} \frac{M_{+-}}{M} \text{Plim} \frac{1}{M_{+-}} \sum^{+-} \frac{\tilde{\alpha}_1 \tilde{\theta} + (\tilde{K}_3 - \tilde{\alpha}_1 \tilde{\alpha}_2 \tilde{K}_0)(y_{m2} - \tilde{\mu}_2)}{\tilde{D}_2}$$

$$+ \text{Plim} \frac{M_{-+}}{M} \text{Plim} \frac{1}{M_{-+}} \sum^{-+} \frac{\tilde{\alpha}_1 \tilde{\theta} - (\tilde{K}_3 - \tilde{\alpha}_1 \tilde{\alpha}_2 \tilde{K}_0)(y_{m2} - \tilde{\mu}_2)}{\tilde{D}_3}$$

$$+ \text{Plim} \frac{M_{--}}{M} \text{Plim} \frac{1}{M_{--}} \sum^{--} \frac{\tilde{\alpha}_1 \tilde{\theta} - (\tilde{K}_3 + \tilde{\alpha}_1 \tilde{\alpha}_2 \tilde{K}_0)(y_{m2} - \tilde{\mu}_2)}{\tilde{D}_4} = 0$$

$$(5.78) \quad \tilde{\alpha}_2 \left[\text{Plim} \frac{M_{2+}}{M} - \frac{\text{Plim}_{2-}}{M} \right]$$

$$- \text{Plim} \frac{M_{++}}{M} \text{Plim} \frac{1}{M_{++}} \sum^{++} \frac{\tilde{\alpha}_2 \tilde{\theta} + (\tilde{K}_3 + \tilde{\alpha}_1 \tilde{\alpha}_2 \tilde{K}_0)(y_{m1} - \tilde{\mu}_1)}{\tilde{D}_1}$$

$$+ \text{Plim} \frac{M_{+-}}{M} \text{Plim} \frac{1}{M_{+-}} \sum^{+-} \frac{\tilde{\alpha}_2 \tilde{\theta} - (\tilde{K}_3 - \tilde{\alpha}_1 \tilde{\alpha}_2 \tilde{K}_0)(y_{m1} - \tilde{\mu}_1)}{\tilde{D}_2}$$

$$- \text{Plim} \frac{M_{-+}}{M} \text{Plim} \frac{1}{M_{-+}} \sum^{-+} \frac{\tilde{\alpha}_2 \tilde{\theta} + (\tilde{K}_3 - \tilde{\alpha}_1 \tilde{\alpha}_2 \tilde{K}_0)(y_{m1} - \tilde{\mu}_1)}{\tilde{D}_3}$$

$$+ \text{Plim} \frac{M_{--}}{M} \text{Plim} \frac{1}{M_{--}} \sum^{--} \frac{\tilde{\alpha}_2 \tilde{\theta} - (\tilde{K}_3 + \tilde{\alpha}_1 \tilde{\alpha}_2 \tilde{K}_0)(y_{m1} - \tilde{\mu}_1)}{\tilde{D}_4} = 0$$

$$(5.79) \quad \frac{1}{\tilde{\alpha}_1} - \text{Plim} \frac{1}{M} \sum_{m=1}^{M} |y_{m1} - \tilde{\mu}_1|$$

$$+ \text{Plim} \frac{1}{M} \sum_{m=1}^{M} \frac{\tilde{\theta}|y_{m1} - \tilde{\mu}_1| + \tilde{\alpha}_2 \tilde{K}_0 |y_{m1} - \tilde{\mu}_1||y_{m2} - \tilde{\mu}_2|}{\tilde{D}_5} = 0$$

$$(5.80) \quad \frac{1}{\tilde{\alpha}_2} - \text{Plim} \frac{1}{M} \sum_{m=1}^{M} |y_{m2} - \tilde{\mu}_2|$$

$$+ \text{Plim} \frac{1}{M} \sum_{m=1}^{M} \frac{\tilde{\theta}|y_{m2} - \tilde{\mu}_2| + \tilde{\alpha}_1 \tilde{K}_0 |y_{m1} - \tilde{\mu}_1||y_{m2} - \tilde{\mu}_2|}{\tilde{D}_5} = 0$$

$$(5.81) \quad -\frac{1}{\tilde{K}_0 + \tilde{\theta}} + \text{Plim} \frac{1}{M} \sum_{m=1}^{M} \frac{1 + \tilde{\alpha}_1 \tilde{\alpha}_2 |y_{m1} - \tilde{\mu}_1||y_{m2} - \tilde{\mu}_2|}{\tilde{D}_5} = 0$$

$$(5.82) \quad -\frac{1}{\tilde{K}_0 + \tilde{\theta}} + \text{Plim} \frac{1}{M} \sum_{m=1}^{M} \frac{\tilde{\alpha}_1 |y_{m1} - \tilde{\mu}_1| + \tilde{\alpha}_2 |y_{m2} - \tilde{\mu}_2|}{\tilde{D}_5} = 0$$

$$(5.83) \quad \text{Plim} \frac{1}{M} \sum_{m=1}^{M} \frac{(y_{m1} - \tilde{\mu}_1)(y_{m2} - \tilde{\mu}_2)}{\tilde{D}_5} = 0$$

Again we have a very messy but (in principal) useful set of equations. To show that μ_i's are consistent , following the same logic as in the univariate case, we only need to show that the above equations are satisfied by $\tilde{\mu}_i = \mu_i$ (and some values for $\tilde{\theta}$, $\tilde{\alpha}_i$, $\tilde{\rho}$, $\tilde{K}_0$, $\tilde{K}_3$).

With $\tilde{\mu}_i = \mu_i$, the second terms in both (5.79) and (5.80) are easy to evaluate, while other terms are very messy. But the actual values of the parameters : $\tilde{\alpha}_i$, $\tilde{\rho}$, $\tilde{\theta}$, $\tilde{K}_0$, $\tilde{K}_3$ are not important as long as acceptable solutions to the above (5.79)-(5.83) exists. By

acceptable we mean that $\tilde{a}_i$, $\tilde{K}_0$, $\tilde{\rho}$, $\tilde{\theta}$, be positive, and $\tilde{K}_3$ be less than $\tilde{a}_1\tilde{a}_2\tilde{K}_0$.

To evaluate (5.77) and (5.78) with $\tilde{\mu}_i = \mu_i$, we note that;

$$P\text{-}lim \left(\frac{M_{1+}}{M} \right) = P\ (y_1 > \mu_1) = P\ (y_1 < \mu_1) = P\left(\frac{M_{1-}}{M} \right)$$

because of the symmetry assumption.

Finally, it is obvious that the two equations (5.77) and (5.78) hold provided that the joint density of the error terms is symmetric around zero and certain expected values exist:

$$(5.84) \quad Plim \ \frac{1}{M_{++}} \sum^{++} \frac{\tilde{\alpha}_1\tilde{\theta}+(\tilde{K}_3+\tilde{\alpha}_1\tilde{\alpha}_2\tilde{K}_0)\epsilon_{m2}}{\tilde{D}_1} = \underset{++}{E} \left\{ \frac{\tilde{\alpha}_1\tilde{\theta}+(\tilde{K}_3+\tilde{\alpha}_1\tilde{\alpha}_2\tilde{K}_0)\epsilon_2}{\tilde{D}_1} \right\}$$

$$= \underset{-}{E} \left\{ \frac{\tilde{\alpha}_1\tilde{\theta}-(\tilde{K}_3+\tilde{\alpha}_1\tilde{\alpha}_2\tilde{K}_0)\epsilon_2}{\tilde{D}_4} \right\} = Plim \ \frac{1}{M_{--}} \sum^{--} \frac{\tilde{\alpha}_1\tilde{\theta}-(\tilde{K}_3+\tilde{\alpha}_1\tilde{\alpha}_2\tilde{K}_0)\epsilon_{m2}}{\tilde{D}_4} \right\}$$

The same can be said about the other terms in equation (5.77), i.e.

$$(5.85) \quad Plim \ \frac{1}{M_{+-}} \sum^{+-} \frac{\tilde{\alpha}_1\tilde{\theta}+(\tilde{K}_3-\tilde{\alpha}_1\tilde{\alpha}_2\tilde{K}_0)\epsilon_{m2}}{\tilde{D}_2} = \underset{+-}{E} \frac{\tilde{\alpha}_1\tilde{\theta}+(\tilde{K}_3-\tilde{\alpha}_1\tilde{\alpha}_2\tilde{K}_0)\epsilon_2}{\tilde{D}_2}$$

$$= \underset{-+}{E} \frac{\tilde{\alpha}_1\tilde{\theta}-(\tilde{K}_3-\tilde{\alpha}_1\tilde{\alpha}_2\tilde{K}_0)\epsilon_2}{\tilde{D}_3} = Plim \ \frac{1}{M_{-+}} \frac{\tilde{\alpha}_1\tilde{\theta}-(\tilde{K}_3-\tilde{\alpha}_1\tilde{\alpha}_2\tilde{K}_0)\epsilon_{m2}}{\tilde{D}_3}$$

The same procedure can be applied to (5.78). Thus all (5.77)-(5.83) are indeed satisfied for $\tilde{\mu}_i = \mu_i$ (and for $\tilde{a}_i$, $\tilde{\theta}$, $\tilde{K}_0$, $\tilde{K}_3$ as being implied by their respective equations (5.79)-(5.83) . The Sargan

"MLE" is consistent for μ_1 and μ_2.

This result can also be obtained in cases involving some regressors other than just a constant term. The proof is basically the same, and the result is that the Sargan "MLE's" are consistent for the coefficients of all of the regressors.

5.5 Conclusions

In this chapter we defined a bivariate Sargan distribution. This was the result of a series of attempts to derive or approximate a bivariate (or multivariate) distribution function with certain properties. Basically, we want a distribution which reasonably approximates the bivariate normal, which has Sargan marginals, and which has an easily computable c.d.f. Our first-order Sargan distribution (5.36) has these properties.

We calculated and compared the densities of the bivariate Sargan and bivariate normal distributions to show that this bivariate Sargan density is very close to the bivariate standard normal. (Of course, it has more parameters than the standard normal.) The agreement between the two distributions is in fact very close except around the turning points (X_1 and X_2 between 1 and 2, say).

Finally, we turned to the question of robustness of inferences based on the Sargan density. Here, we looked only at a rather simple model, a seemingly unrelated regression model , with only a constant term in each equation. Of course the MLE's based on a normality assumption are consistent, even if the errors are not normal. The more

interesting question is the effect on the estimates of assuming Sargan where the true distribution is not Sargan. Here we showed that the Sargan "MLE's" are still asymptotically unbiased, so long as the true distribution of the errors is symmetric around zero. This result can also be obtained in regressions with multiple regressors.

For more complicated models, the Sargan MLE's will not always be consistent if the errors do not actually have a Sargan distribution. As we found for the univariate case in chapters 3 and 4, the extent of the asymptotic bias must be model dependent. But the above result at least is the basis for the expectation that the bias will depend strongly on the extent of censoring or truncation in the model: with complete observations, there is no bias.

TABLE 5.1

Comparison of Bivariate Sargan and Normal Densities
(Normal = top entry, Sargan = bottom entry)

$\rho = -.15$

X_1 \ X_2	3.0	2.5	2.0	1.5	1.0	.5	0
-3.0	0.0						
	0.0						
-2.5	0.0	.001					
	0.0	0.0					
-2.0	.001	.002	.005				
	.001	.001	.003				
-1.5	.001	.004	.010	.023			
	.001	.003	.006	.012			
-1.0	.002	.006	.017	.038	.067		
	.002	.005	.012	.025	.051		
-.5	.002	.007	.021	.050	.092	.130	
	.004	.010	.021	.045	.090	.156	
0.0	.002	.007	.021	.051	.097	.142	.161
	.005	.013	.028	.059	.114	.184	.161
.5	.001	.005	.016	.040	.079	.120	
	.002	.006	.013	.028	.060	.115	
1.0	.001	.003	.009	.024	.050		
	.001	,002	.006	.013	.029		
1.5	0.0	.001	.004	.011			
	0.0	.001	.002	.006			
2.0	0.0	0.0	.001				
	0.0	0.0	.001				
2.5	0.0	0.0					
	0.0	0.0					
3.0	0.0						
	0.0						

TABLE 5.2

Comparison of Bivariate Sargan and Normal Densities
(Normal = top entry, Sargan = bottom entry)

$$\rho = -.10$$

X_1 \\ X_2	3.0	2.5	2.0	1.5	1.0	.5	0
-3.0	0.0						
	0.0						
-2.5	0.0	0.0					
	0.0	0.0					
-2.0	0.0	.001	.004				
	0.0	.001	.002				
-1.5	.001	.003	.010	.021			
	.001	.002	.006	.011			
-1.0	.001	.005	.016	.036	.064		
	.002	.005	.011	.023	.047		
-.5	.002	.007	.021	.049	.089	.127	
	.004	.009	.020	.042	.085	.149	
0.0	.002	.007	.021	.051	.097	.141	.160
	.005	.013	.028	.059	.114	.184	.160
.5	.001	.005	.017	.042	.081	.121	
	.003	.006	.014	.031	.065	.122	
1.0	.001	.003	.010	.027	.053		
	.001	.003	.007	.015	.032		
1.5	0.0	.001	.005	.013			
	.001	.001	.003	.007			
2.0	0.0	.001	.002				
	0.0	.001	.001				
2.5	0.0	0.0					
	0.0	0.0					
3.0	0.0						
	0.0						

TABLE 5.3

Comparison of Bivariate Sargan and Normal Densities
(Normal = top entry, Sargan = bottom entry)

$\rho = -.05$

X_1 \ X_2	3.0	2.5	2.0	1.5	1.0	.5	0
-3.0	0.0						
	0.0						
-2.5	0.0	0.0					
	0.0	0.0					
-2.0	0.0	.001	.004				
	0.0	.001	.002				
-1.5	.001	.003	.008	.019			
	.001	.002	.005	.010			
-1.0	.001	.005	.014	.034	.061		
	.002	.004	.010	.021	.043		
-.5	.002	.007	.020	.047	.087	.126	
	.003	.008	.018	.040	.080	.142	
0.0	.002	.007	.021	.052	.097	.141	.159
	.005	.013	.028	.059	.114	.184	.159
.5	.001	.006	.018	.044	.083	.122	
	.003	.007	.015	.034	.070	.129	
1.0	.001	.004	.012	.029	.056		
	.001	.003	.008	.017	.036		
1.5	0.0	.002	.006	.015			
	.001	.002	.004	.008			
2.0	0.0	.001	.002				
	0.0	.001	.002				
2.5	0.0	0.0					
	0.0	0.0					
3.0	0.0						
	0.0						

TABLE 5.4

Comparison of Bivariate Sargan and Normal Densities
(Normal = top entry, Sargan = bottom entry)

$\rho = 0.0$

X_1 \ X_2	3.0	2.5	2.0	1.5	1.0	.5	0
-3.0	0.0						
	0.0						
-2.5	0.0	0.0					
	0.0	0.0					
-2.0	0.0	.001	.003				
	0.0	.001	.002				
-1.5	.001	.002	.007	.017			
	.001	.002	.004	.009			
-1.0	.001	.004	.013	.031	.059		
	.002	.004	.009	.019	.039		
-.5	.002	.006	.019	.046	.085	.124	
	.003	.007	.017	.037	.075	.135	
0.0	.002	.007	.022	.052	.097	.141	.159
	.005	.013	.028	.059	.114	.184	.159
.5	.002	.006	.019	.046	.085	.124	
	.003	.007	.017	.037	.075	.135	
1.0	.001	.004	.013	.031	.059		
	.002	.004	.009	.019	.040		
1.5	.001	.002	.007	.017			
	.001	.002	.004	.010			
2.0	0.0	.001	.003				
	0.0	.001	.002				
2.5	0.0	0.0					
	0.0	0.0					
3.0	0.0						
	0.0						

TABLE 5.5

Comparison of Bivariate Sargan and Normal Densities
(Normal = top entry, Sargan = bottom entry)

$\rho = .05$

X_1 \\ X_2	3.0	2.5	2.0	1.5	1.0	.5	0
-3.0	0.0						
	0.0						
-2.5	0.0	0.0					
	0.0	0.0					
-2.0	0.0	.001	.002				
	0.0	.001	.002				
-1.5	0.0	.002	.006	.015			
	.001	.002	.004	.008			
-1.0	.001	.004	.014	.029	.056		
	.001	.003	.007	.017	.036		
-.5	.001	.006	.018	.044	.083	.122	
	.003	.007	.015	.034	.070	.129	
0.0	.002	.007	.021	.052	.097	.141	.159
	.005	.013	.028	.059	.114	.184	.159
.5	.002	.007	.020	.047	.087	.126	
	.003	.008	.018	.039	.080	.142	
1.0	.001	.005	.014	.034	.061		
	.002	.004	.010	.021	.043		
1.5	.001	.003	.008	.019			
	.001	.002	.005	.010			
2.0	0.0	.001	.004				
	0.0	.001	.002				
2.5	0.0	0.0					
	0.0	0.0					
3.0	0.0						
	0.0						

TABLE 5.6

Comparison of Bivariate Sargan and Normal Densities
(Normal = top entry, Sargan = bottom entry)

ρ = .10

X_1 \ X_2	3.0	2.5	2.0	1.5	1.0	.5	0
-3.0	0.0						
	0.0						
-2.5	0.0	0.0					
	0.0	0.0					
-2.0	0.0	.001	.002				
	0.0	.001	.001				
-1.5	0.0	.001	.005	.013			
	.001	.001	.003	.007			
-1.0	.001	.003	.010	.027	.053		
	.001	.003	.007	.015	.032		
-.5	.001	.005	.017	.042	.081	.121	
	.002	.006	.014	.031	.065	.115	
0.0	.002	.007	.021	.051	.097	.141	.160
	.005	.013	.028	.059	.114	.184	.160
.5	.002	.007	.021	.049	.089	.128	
	.004	.009	.020	.042	.085	.149	
1.0	.001	.005	.016	.036	.067		
	.002	.005	.011	.023	.047		
1.5	.001	.003	.009	.023			
	.001	.002	.005	.011			
2.0	0.0	.001	.005				
	0.0	.001	.002				
2.5	0.0	.001					
	0.0	0.0					
3.0	0.0						
	0.0						

TABLE 5.7

Comparison of Bivariate Sargan and Normal Densities
(Normal = top entry, Sargan = bottom entry)

$\rho = .15$

X_1 \ X_2	3.0	2.5	2.0	1.5	1.0	.5	0
-3.0	0.0						
	0.0						
-2.5	0.0	0.0					
	0.0	0.0					
-2.0	0.0	0.0	.001				
	0.0	0.0	.001				
-1.5	0.0	.001	.004	.011			
	0.0	.001	.002	.006			
-1.0	.001	.003	.009	.024	.050		
	.001	.002	.006	.013	.029		
-.5	.001	.005	.016	.040	.079	.120	
	.002	.006	.013	.028	.060	.115	
0.0	.002	.007	.021	.051	.097	.142	.161
	.005	.013	.028	.059	.114	.184	.161
.5	.002	.007	.021	.050	.092	.130	
	.004	.010	.021	.045	.090	.156	
1.0	.002	.006	.017	.038	.067		
	.002	.005	.012	.025	.051		
1.5	.001	.004	.011	.023			
	.001	.003	.006	.012			
2.0	.001	.002	.005				
	.001	.001	.003				
2.5	0.0	.001					
	0.0	.001					
3.0	0.0						
	0.0						

DERIVATION OF THE COVARIANCES

Here, we use the following relationship:

$$\int_0^\infty x^n e^{-\alpha x}\, dx = (-1)^n \int_{-\infty}^0 y^n e^{\alpha y}\, dy = \frac{n!}{\alpha^{n+1}} \qquad n > 0 \qquad (A.1)$$

$$COV(X,Y) = E(X,Y) = \int_{-\infty}^\infty \int_{-\infty}^\infty xy f(x,y)\,dx\,dy \qquad (A.2)$$

$f(x,y)$ is defined as in (5.14)

$$I_1 = \int_{-\infty}^0 \int_{-\infty}^0 xy\, e^{\alpha_1 x + \alpha_2 y} (K_0 - K_1 x - K_2 y + K_3 xy)\,dx\,dy$$

$$= \int_{-\infty}^0 y e^{\alpha_2 y} \left[-\frac{K_0 - K_2 y}{\alpha_1^2} - 2\frac{K_1 - K_3 y}{\alpha_2^3} \right] dy = \frac{1}{\alpha_1^2}\left\{ \frac{K_0 + \frac{2K_1}{\alpha_1}}{\alpha_2^2} + 2\frac{K_2 + \frac{2K_3}{\alpha_1}}{\alpha_2^3} \right\}$$

$$= \frac{1}{\alpha_1^3 \alpha_2^3} \left\{ \alpha_1 \alpha_2 K_0 + 2\alpha_2 K_1 + 2\alpha_1 K_2 + 4K_3 \right\} \qquad (A-3)$$

$$I_2 = \int_{-\infty}^0 y e^{\alpha_2 y} \int_0^\infty x e^{-\alpha_2 x} (K_0 + K_1 x - K_2 y - K_3 xy)\,dx\,dy$$

$$= \int_{-\infty}^0 y e^{\alpha_2 y} \left\{ \frac{K_0}{\alpha_1^2} + \frac{2K_1}{\alpha_1^3} - \frac{K_2 y}{\alpha_1^2} - \frac{2K_3 y}{\alpha_1^3} \right\} dy$$

$$= \frac{-1}{\alpha_1^3 \alpha_2^3} (\alpha_1 \alpha_2 K_0 + 2\alpha_2 K_1 + 2\alpha_1 K_2 + 4K_3) \qquad (A-4)$$

$$I_3 = \int_0^\infty e^{-\alpha_2 y}\, y \int_{-\infty}^0 x e^{\alpha_2 x} (K_0 - K_1 x + K_2 y - K_3 xy)\,dx\,dy$$

$$= \int_0^\infty y e^{-\alpha_2 y} \left\{ -\frac{K_0}{\alpha_1^2} - \frac{2K_1}{\alpha_1^3} - \frac{K_2 y}{\alpha_1^2} - \frac{2K_3 y}{\alpha_1^3} \right\} dy$$

$$= \frac{-1}{\alpha_1^2} \left\{ \left(K_0 + \frac{2K_1}{\alpha_1} \right) \frac{1}{\alpha_2^2} + 2 \frac{K_2 + 2K_3/\alpha_1}{\alpha_2^3} \right\}$$

$$= \frac{-1}{\alpha_1^3 \alpha_2^3} (\alpha_1 \alpha_2 K_0 + 2\alpha_2 K_1 + 2\alpha_1 K_2 + 4K_3) \qquad (A.5)$$

$$I_4 = \int_0^\infty e^{-\alpha_2 y} y \int_0^\infty e^{-\alpha_1 x} (K_0 x + K_1 x^2 + K_2 xy + K_3 x^2 y) dx\, dy \qquad (A.6)$$

$$= \int_0^\infty y e^{-\alpha_2 y} \left\{ \frac{K_0}{\alpha_1^2} + \frac{2K_1}{\alpha_1^3} + \frac{K_2 y}{\alpha_1^2} + \frac{2K_3 y}{\alpha_1^3} \right\} dy$$

$$= \frac{1}{\alpha_2^2} \left\{ \frac{K_0}{\alpha_2^2} + \frac{2K_1}{\alpha_1 \alpha_2^2} + \frac{2K_2}{\alpha_2^3} + \frac{4K_3}{\alpha_1 \alpha_2^3} \right\}$$

$$= \frac{1}{\alpha_1^3 \alpha_2^3} \left\{ \alpha_1 \alpha_2 K_0 + 2\alpha_2 K_1 + 2\alpha_1 K_2 + 4K_3 \right\}$$

Therefore

$$COV(X,Y) = 0 \qquad (A.7)$$

The same process is applied to $f(x_1, x_2)$ as defined in (5.27):

That is:

$$R_1 = \int_{-\infty}^{0} \int_{-\infty}^{0} x_1 x_2 e^{\alpha_1 x_1 + \alpha_2 x_2} (-\theta\alpha_1 x_1 - \theta\alpha_2 x_2 + K_3 x_1 x_2) dx_1 dx_2$$

$$= \int_{-\infty}^{0} x_2 e^{+\alpha_2 x_2} [- \frac{2\alpha_1 \theta}{\alpha_1^3} + \frac{\alpha_2 \theta x_2}{\alpha_1^2} + \frac{2K_3 x_2}{\alpha_1^3}] \, dx_2$$

$$= \frac{2\theta}{\alpha_1^2 \alpha_2^2} + \frac{2\theta\alpha_2}{\alpha_1^2 \alpha_2^3} + \frac{4K_3}{\alpha_1^3 \alpha_2^3} = \frac{4}{\alpha_1^3 \alpha_2^3} (\alpha_1 \alpha_2 \theta + K_3) \qquad (A.8)$$

and similarly for other intervals; R_2, R_3, R_4, we get:

$$R_2 = \int_{-\infty}^{0} \int_{0}^{\infty} x_1 x_2 e^{-\alpha_1 x_1 + \alpha_2 x_2} (\theta\alpha_1 x_1 - \theta\alpha_2 x_2 + K_3 x_1 x_2) dx_1 dx_2$$

$$= \int_{-\infty}^{0} x_2 e^{\alpha_2 x_2} (\frac{2\theta x_1}{\alpha_1^3} - \frac{\theta\alpha_2 x_2}{\alpha_1^2} + \frac{2K_3 x_2}{\alpha_1^3}) \, dx_2$$

$$= \frac{-2\theta}{\alpha_1^2 \alpha_2^2} - \frac{2\theta}{\alpha_1^2 \alpha_2^2} + \frac{4K_3}{\alpha_1^3 \alpha_2^3} = \frac{-4}{\alpha_1^3 \alpha_2^3} (\alpha_1 \alpha_2 \theta - K_3) \qquad (A.9)$$

$$R_3 = \frac{-4}{\alpha_1^3 \alpha_2^3} (\alpha_1 \alpha_2 \theta - K_3) \tag{A.10}$$

$$R_4 = \frac{4}{\alpha_1^3 \alpha_2^3} (\alpha_1 \alpha_2 \theta + K_3) \tag{A.11}$$

Therefore,

$$COV(X_1, X_2) = E(X_1, X_2) = \frac{\alpha_1 \alpha_2}{8\theta} (R_1 + R_2 + R_3 + R_4) =$$

$$\frac{2K_3}{\theta \alpha_1^2 \alpha_2^2} \tag{A.12}$$

DERIVATION OF MOMENT GENERATING FUNCTION AND CUMULATIVE

DISTRIBUTION FUNCTION OF BIVARIATE SARGAN

i) Moment generating function:

let

$$f(x,y) = \frac{\alpha_1 \alpha_2}{8(K_0+\theta)} e^{-\alpha_1|x_1|-\alpha_2|x_2|} (K_0+\alpha_1\theta|x_1|+\alpha_2\theta|x_2|$$

$$+K_3 x_1 x_2 + \alpha_1 \alpha_2 K_0 |x_1||x_2| \qquad (B.1)$$

as defined in (5.36)

$$M(T) = E(e^{t_1 x_1 + t_2 x_2})$$

$$B_1 = \int_{-\infty}^{0}\int_{-\infty}^{0} e^{(\alpha_1+t_1)x_1+(\alpha_2+t_2)x_2} (K_0+\alpha_1\theta x_1-\alpha_2\theta x_2+(K_3+\alpha_1\alpha_2 K_0)x_1 x_2)dx_1 dx_2$$

$$= \int_{-\infty}^{0} e^{(\alpha_2+t_2)x_2} \{\frac{K_0-\alpha_2\theta x_2}{\alpha_1+t_1} + \frac{\alpha_1\theta-(K_3+\alpha_1\alpha_2 K_0)x_2}{(\alpha_1+t_1)^2}\} dx_2$$

$$= \frac{(\alpha_1+t_1)K_0+\alpha_1\theta}{(\alpha_1+t_1)^2(\alpha_2+t_2)} + \frac{(\alpha_1+t_1)\alpha_2\theta+(K_3+\alpha_1\alpha_2 K_0)}{(\alpha_1+t_1)^2(\alpha_2+t_2)^2} \qquad (B.2)$$

$$B_2 = \int_{-\infty}^{0} \int_{0}^{\infty} e^{-(\alpha_1-t_1)x_1+(\alpha_2+t_2)x_2}(K_0+\alpha_1\theta x_1-\alpha_2\theta x_2+(K_3-\alpha_1\alpha_2 K_0)x_1 x_2)dx_1 dx_2$$

$$= \int_{-\infty}^{0} e^{(\alpha_2+t_2)x_2}\{\frac{K_0-\alpha_2\theta x_2}{\alpha_1-t_1} + \frac{\alpha_1\theta+(K_3-\alpha_1\alpha_2 K_0)x_2}{(\alpha_1-t_1)^2}\}\ dx_2$$

$$= \frac{(\alpha_1-t_1)K_0+\alpha_1\theta}{(\alpha_1-t_1)^2(\alpha_2+t_2)} + \frac{(\alpha_1-t_1)\alpha_2\theta-(K_3-\alpha_1\alpha_2 K_0)}{(\alpha_1-t_1)^2(\alpha_2+t_2)^2} \qquad (B.3)$$

$$B_3 = \int_{0}^{\infty} \int_{-\infty}^{0} e^{(\alpha_1+t_1)x_1-(\alpha_2-t_2)x_2}[K_0-\alpha_1\theta x_1+\alpha_2\theta x_2+(K_3-\alpha_1\alpha_2 K_0)x_1 x_2]dx_1 dx_2$$

$$= \int_{0}^{\infty} e^{-(\alpha_2-t_2)x_2}\{\frac{K_0+\alpha_2\theta x_2}{\alpha_1+t_1} + \frac{\alpha_1\theta-(K_3-\alpha_1\alpha_2 K_0)x_2}{(\alpha_1+t_1)^2}\}dx_2$$

$$= \frac{(\alpha_1+t_1)K_0+\alpha_1\theta}{(\alpha_1+t_1)^2(\alpha_2-t_2)} + \frac{-(K_3-\alpha_1\alpha_2 K_0)+\alpha_2\theta(\alpha_1+t_1)}{(\alpha_1+t_1)^2(\alpha_2-t_2)^2} \qquad (B.4)$$

$$B_4 = \int_{0}^{\infty} \int_{0}^{\infty} e^{-(\alpha_1-t_1)x_1-(\alpha_2-t_2)x_2}[K_0+\alpha_1\theta x_1+\alpha_2\theta x_2+(K_3+\alpha_1\alpha_2 K_0)x_1 x_2]dx_1 dx_2$$

$$= \int_0^\infty \left\{ \frac{K_0 + \alpha_2 \theta x_2}{\alpha_1 - t_1} + \frac{\alpha_1 \theta + (K_3 + \alpha_1 \alpha_2 K_0) x_2}{(\alpha_1 - t_1)^2} \right\} e^{-(\alpha_2 - t_2)x_2} dx_2$$

$$= \frac{(\alpha_1 - t_1)K_0 + \alpha_1 \theta}{(\alpha_1 - t_1)^2 (\alpha_2 - t_2)} + \frac{(\alpha_1 - t_1)\alpha_2 \theta + (K_3 + \alpha_1 \alpha_2 K_0)}{(\alpha_1 - t_1)^2 (\alpha_2 - t_2)^2} \tag{B.5}$$

$$M(T) = \frac{\alpha_1 \alpha_2}{8(K_0 + \theta)} \{ B_1 + B_2 + B_3 + B_4 \}$$

$$= \frac{\alpha_1 \alpha_2}{8(K_0 + \theta)} \left\{ \frac{(\alpha_2 + t_2)[(\alpha_1 + t_1)K_0 + \alpha_1 \theta] + \alpha_2 \theta(\alpha_1 + t_1) + K_3 + \alpha_1 \alpha_2 K_0}{(\alpha_1 + t_1)^2 (\alpha_2 + t_2)^2} \right.$$

$$+ \frac{(\alpha_2 + t_2)[(\alpha_1 - t_1)K_0 + \alpha_1 \theta] + (\alpha_1 - t_1)\alpha_2 \theta - (K_3 - \alpha_1 \alpha_2 K_0)}{(\alpha_1 - t_1)^2 (\alpha_2 + t_2)^2}$$

$$+ \frac{(\alpha_2 - t_2)[(\alpha_1 + t_1)K_0 + \alpha_1 \theta] - (K_3 - \alpha_1 \alpha_2 K_0) + \alpha_2 \theta(\alpha_1 + t_1)}{(\alpha_1 + t_1)^2 (\alpha_2 - t_2)^2}$$

$$+ \left. \frac{(\alpha_2 - t_2)[(\alpha_1 - t_1)K_0 + \alpha_1 \theta] + \alpha_2 \theta(\alpha_1 - t_1) + (K_3 + \alpha_1 \alpha_2 K_0)}{(\alpha_1 - t_1)^2 (\alpha_2 - t_2)^2} \right\}$$

$$= \left\{ \frac{K_0}{(\alpha_1+t_1)(\alpha_2+t_2)} + \frac{\alpha_1\theta}{(\alpha_1+t_1)^2(\alpha_2+t_2)} + \frac{\alpha_2\theta}{(\alpha_1+t_1)(\alpha_2+t_2)^2} \right.$$

$$+ \frac{\alpha_1\alpha_2K_0+K_3}{(\alpha_1+t_1)^2(\alpha_2+t_2)^2}$$

$$\frac{K_0}{(\alpha_1-t_1)(\alpha_2+t_2)} + \frac{\alpha_1\theta}{(\alpha_1+t_1)^2(\alpha_2+t_2)} + \frac{\alpha_2\theta}{(\alpha_1-t_1)(\alpha_2+t_2)^2}$$

$$+ \frac{\alpha_1\alpha_2K_0-K_3}{(\alpha_1-t_1)^2(\alpha_2+t_2)^2} \qquad\qquad \text{(B.6)}$$

$$\frac{K_0}{(\alpha_1+t_1)(\alpha_2-t_2)} + \frac{\alpha_1\theta}{(\alpha_1+t_1)^2(\alpha_2-t_2)} + \frac{\alpha_2\theta}{(\alpha_1+t_1)(\alpha_2-t_2)^2}$$

$$+ \frac{\alpha_1\alpha_2K_0-K_3}{(\alpha_1+t_1)^2(\alpha_2-t_2)^2}$$

$$\frac{K_0}{(\alpha_1-t_1)(\alpha_2-t_2)} + \frac{\alpha_1\theta}{(\alpha_1-t_1)^2(\alpha_2-t_2)} + \frac{\alpha_2\theta}{(\alpha_1-t_1)(\alpha_2-t_2)^2}$$

$$\left. + \frac{\alpha_1\alpha_2K_0+K_3}{(\alpha_1-t_1)^2(\alpha_2-t_2)^2} \right\} \; \frac{\alpha_1\alpha_2}{8(K_0+\theta)}$$

Since,

$$\frac{\partial^{n+m}}{\partial X_1^n \partial X_2^m} \left(\frac{1}{X_1 X_2}\right) = \frac{(-1)^{n+m} \, n! m!}{X_1^{n+1} \, X_2^{m+1}} \quad \text{or in general} \tag{B.7}$$

$$\frac{\partial^{n+m}}{\partial X_1^n \partial X_2^m} \left(\frac{1}{X^i X^j}\right) = \frac{(-1)^{n+m} (n-i-1)!(m+j-1)!}{X_1^{n+1} X_2^{m+j} (i-1)!(j-1)!} \tag{B.8}$$

$$\frac{\partial^{n+m}}{\partial t_1^n \partial t_2^m} [M(t)] = \frac{\alpha_1 \alpha_2}{8(K_0+\theta)} \left\{ \frac{(-1)^{n+m} \, n! \, m! \, K_0}{(\alpha_1+t_1)^{n+1} (\alpha_2+t_2)^{m+1}} \right.$$

$$\frac{(-1)^{n+m} (n+1)! \, m! \, \alpha_1 \theta}{(\alpha_1+t_1)^{n+2}(\alpha_2+t_2)^{m+1}}$$

$$+ \frac{(-1)^{n+m} \, n! \, (m+1)! \, \alpha_2 \theta}{(\alpha_1+t_1)^{n+1}(\alpha_2+t_2)^{m+2}} + \frac{(-1)^{n+m} (n+1)! \, (m+1)! \, (\alpha_1 \alpha_2 K_0 + K_3)}{(\alpha_1+t_1)^{n+2}(\alpha_2+t_2)^{m+2}}$$

$$+ \frac{(-1)^m \, n! \, m! \, K_0}{(\alpha_1-t_1)^{n+1}(\alpha_2+t_2)^{m+1}} + \frac{(-1)^m (n+1)! \, m! \, \alpha_1 \theta}{(\alpha_1-t_1)^{n+2}(\alpha_2+t_2)^{m+1}}$$

$$\frac{(-1)^m \, n! \, (m+1)! \, \alpha_2 \theta}{(\alpha_1 - t_1)^{n+1} (\alpha_2 + t_2)^{m+2}} + \frac{(-1)^m \, (n+1)! \, (m+1)! \, (\alpha_1 \alpha_2 K_0 - K_3)}{(\alpha_1 - t_1)^{n+2} (\alpha_2 + t_2)^{m+2}}$$

$$+ \frac{(-1)^n \, n! \, m! \, K_0}{(\alpha_1 + t_1)^{n+1} (\alpha_2 - t_2)^{m+1}} + \frac{(-1)^n \, (n+1)! \, m! \, \alpha_1 \theta}{(\alpha_1 + t_1)^{n+2} (\alpha_2 - t_2)^{m+1}}$$

$$+ \frac{(-1)^n n! \, (m+1)! \, \alpha_2 \theta}{(\alpha_1 + t_1)^{n+1} (\alpha_2 - t_2)^{m+2}} + \frac{(-1)^n (n+1)! \, (m+1)! \, (\alpha_1 \alpha_2 K_0 - K_3)}{(\alpha_1 + t_1)^{n+2} (\alpha_2 - t_2)^{m+2}}$$

$$+ \frac{(-1)^{2(n+m)} n! \, m! \, K_0}{(\alpha_1 - t_1)^{n+1} (\alpha_2 - t_2)^{m+1}} + \frac{(-1)^{2(n+m)} (n+1)! \, m! \, \alpha_1 \theta}{(\alpha_1 - t_1)^{n+2} (\alpha_2 - t_2)^{m+1}}$$

$$+ \frac{(-1)^{2(n+m)} n! \, (m+1)! \, \alpha_2 \theta}{(\alpha_1 - t_1)^{n+1} (\alpha_2 - t_2)^{m+2}} + \frac{(-1)^{2(n+m)} (n+1)! \, (m+1)! \, (\alpha_1 \alpha_2 K_0 + K_3)}{(\alpha_1 - t_1)^{n+2} (\alpha_2 - t_2)^{m+2}} \}$$

$$\mu_{nm} = \frac{\partial^{n+m} M(t=0)}{\partial t_1^n \partial t_2^m} = \frac{\alpha_1 \alpha_2}{8(K_0 + \theta)} \left\{ \frac{n! \, m! K_0 [1 + (-1)^n + (-1)^m + (-1)^{n+m}]}{\alpha_1^{n+1} \alpha_2^{m+1}} \right.$$

$$\frac{+[1+(-1)^n+(-1)^m+(-1)^{n+m}](n+1)!m!\alpha_1\theta}{\alpha_1^{n+2}\alpha_2^{m+1}}$$

$$+\frac{[1+(-1)^n+(-1)^m+(-1)^{n+m}]n!(m+1)!\alpha_2\theta}{\alpha_1^{n+1}\alpha_2^{m+2}}$$

$$+(n+1)!(m+1)!\frac{[(-1)^{n+m}+(-1)^m+(-1)^n+(-1)^{2(n+m)}]\alpha_1\alpha_2 K_0}{\alpha_1^{n+2}\alpha_2^{m+2}}$$

$$+\frac{[(-1)^{n+m}-(-1)^m-(-1)^n+(-1)^{2(n+m)}]K_3(n+1)!(m+1)!}{\alpha_1^{n+2}\alpha_2^{m+2}}\}$$

Therefore;

$$\mu_{nm} = \frac{n!\ m!\ K_0+(n+1)!\ m!\ \theta+n!\ (m+1)!\ \theta+(n+1)!\ (m+1)!K_0}{2\alpha_1^n\alpha_2^m(K_0+\theta)}$$

if both n, m even

$$= \frac{[n!m!+(n+1)!(m+1)!]K_0+[(n+1)!m!+n!(m+1)!]\theta}{2\alpha_1^n\alpha_2^m(K_0+\theta)} \tag{B.11}$$

$$\mu_{nm} = \frac{(n+1)!(m+1)!K_3}{2\alpha_1^{n+1}\alpha_2^{m+1}(K_0+\theta)} \qquad \text{If both n,m are } \underline{\text{odd}} \qquad \text{(B.12)}$$

$$\mu_{nm} = 0 \qquad \qquad \text{otherwise} \qquad \text{(B.13)}$$

ii) Derivation of Cumulative Distribution Function

1) $\underline{z_1, z_2 > 0}$

$$I_1 = \int_{-\infty}^{0}\int_{-\infty}^{0} e^{\alpha_1 x_1 + \alpha_2 x_2}[K_0 - \alpha_1\theta x_1 - \alpha_2\theta x_2 + (K_3 + \alpha_1\alpha_2 K_0)x_1 x_2]dx_1 dx_2$$

$$= \int_{-\infty}^{0} e^{\alpha_2 x_2}[\frac{K_0 - \alpha_2\theta x_2}{\alpha_1} + \frac{\alpha_1\theta - (K_3 + \alpha_1\alpha_2 K_0)x_2}{\alpha_1^2}] \, dx_2$$

$$= \frac{2(K_0+\theta)\alpha_1\alpha_2 + K_3}{\alpha_1^2\alpha_2^2} \qquad \text{(B.14)}$$

$$I_2 = \int_{-\infty}^{0}\int_{0}^{z_1} e^{-\alpha_1 x_1 + \alpha_2 x_2}[K_0 + \alpha_1\theta x_1 - \alpha_2\theta x_2 + (K_3 - \alpha_1\alpha_2 K_0)x_1 x_2]dx_1 dx_2$$

$$= \int_{-\infty}^{0} - \frac{1}{\alpha_1} \{e^{-\alpha_1 z_1}[K_0+\theta+\alpha_1\theta z_1-\alpha_2\theta x_2+(K_3-\alpha_1\alpha_2 K_0)z_1 x_2$$

$$+ \frac{(K_3-\alpha_1\alpha_2 K_0)x_2}{\alpha_1}]\, e^{\alpha_2 x_2}$$

$$- [K_0+\theta-\alpha_2\theta x_2 + \frac{(K_3-\alpha_1\alpha_2 K_0)x_2}{\alpha_1}]\, e^{\alpha_2 x_2}\}\, dx_2$$

$$= -\frac{1}{\alpha_1} e^{-\alpha_1 z_1}[\frac{K_0+\theta+\alpha_1\theta z_1}{\alpha_2} + \frac{\alpha_2\theta-(K_3-\alpha_1\alpha_2 K_0)z_1}{\alpha_2^2} - \frac{K_3-\alpha_1\alpha_2 K_0}{\alpha_1\alpha_2^2}]$$

$$+ \frac{1}{\alpha_1}[\frac{K_0+\theta}{\alpha_2} + \frac{\alpha_2\theta}{\alpha_2^2} - \frac{K_3-\alpha_1\alpha_2 K_0}{\alpha_1\alpha_2^2}]$$

$$= \frac{-e^{-\alpha_1 z_1}}{\alpha_1^2\alpha_2^2}\{2\alpha_1\alpha_2(K_0+\theta)-K_3+\alpha_1[\alpha_1\alpha_2(K_0+\theta)-K_3]z_1\} + \frac{2\alpha_1\alpha_2(K_0+\theta)-K_3}{\alpha_1^2\alpha_2^2}$$

$$(B.15)$$

$$I_3 = \int_0^{z_2}\int_{-\infty}^0 e^{\alpha_1 x_1-\alpha_2 x_2}[K_0-\alpha_1\theta x_1+\alpha_2\theta x_2+(K_3-\alpha_1\alpha_2 K_0)x_1 x_2]dx_1 dx_2$$

$$= \int_0^{z_2} e^{-\alpha_2 x_2}[\frac{K_0+\alpha_2\theta x_2}{\alpha_1} + \frac{\theta}{\alpha_1} - \frac{(K_3-\alpha_1\alpha_2 K_0)x_2}{\alpha_1^2}]\, dx_2$$

$$= \frac{-e^{-\alpha_2 z_2}}{\alpha_1^2\alpha_2^2}\{2\alpha_1\alpha_2(K_0+\theta)-K_3+\alpha_2[\alpha_1\alpha_2(K_0+\theta)-K_3]z_2\} + \frac{2\alpha_1\alpha_2(K_0+\theta)-K_3}{\alpha_1^2\alpha_2^2}$$

$$(B.16)$$

$$I_4 = \int_0^{z_2} \int_0^{z_1} e^{-\alpha_1 x_1 - \alpha_2 x_2} [K_0 + \alpha_1 \theta x_1 + \alpha_2 \theta x_2 + (K_3 + \alpha_1 \alpha_2 K_0) x_1 x_2] dx_1 dx_2$$

$$= \int_0^{z_2} -\frac{1}{\alpha_1} \{ e^{-\alpha_1 z_1} [K_0 + \theta + \alpha_1 \theta z_1 + \alpha_2 \theta x_2 + (K_3 + \alpha_1 \alpha_2 K_0) z_1 x_2$$

$$+ \frac{(K_3 + \alpha_1 \alpha_2 K_0) x_2}{\alpha_1}]$$

$$- [K_0 + \theta + \alpha_2 \theta x_2 + \frac{(K_3 - \alpha_1 \alpha_2 K_0) x_2}{\alpha_1}] \} \, e^{\alpha_2 x_2} dx_2$$

$$= \frac{e^{-\alpha_1 z_1 - \alpha_2 z_2}}{\alpha_1^2 \alpha_2^2} \{ 2\alpha_1 \alpha_2 (K_0 + \theta) + K_3 + \alpha_1 [\alpha_1 \alpha_2 (K_0 + \theta) + K_3] z_1$$

$$+ \alpha_2 [\alpha_1 \alpha_2 (K_0 + \theta) + K_3] z_2 + \alpha_1 \alpha_2 (K_3 + \alpha_1 \alpha_2 K_0) z_1 z_2 \}$$

$$\frac{-e^{-\alpha_1 z_1}}{\alpha_1^2 \alpha_2^2} \{ 2\alpha_1 \alpha_2 (K_0 + \theta) + K_3 + \alpha_1 [\alpha_1 \alpha_2 (K_0 + \theta) + K_3] z_1 \}$$

$$\frac{-e^{-\alpha_2 z_2}}{\alpha_1^2 \alpha_2^2} \{ 2\alpha_1 \alpha_2 (K_0 + \theta) + K_3 + \alpha_2 [\alpha_1 \alpha_2 (K_0 + \theta) + K_3] z_2 \}$$

$$+ \frac{1}{\alpha_1^2 \alpha_2^2} \{ 2\alpha_1 \alpha_2 (K_0 + \theta) + K_3 \}$$

$$= \frac{1}{\alpha_1^2 \alpha_2^2} \{ e^{-\alpha_1 z_1 - \alpha_2 z_2} [w_1 + \alpha_1 w_2 z_1 + \alpha_2 w_2 z_2 + w_3 z_1 z_2]$$

$$-e^{-\alpha_1 z_1} [w_1 + \alpha_1 w_2 z_1] - e^{-\alpha_2 z_2} [w_1 + \alpha_2 w_2 z_2] + w_1 \} \qquad \text{(B.17)}$$

$$F(z_1 z_2) = A(I_1 + I_2 + I_3 + I_4)$$

$$= \frac{A}{\alpha_1^2 \alpha_2^2} \{ e^{-\alpha_1 z_1 - \alpha_2 z_2} [w_1 + \alpha_1 w_2 z_1 + \alpha_2 w_2 z_2 + w_3 z_1 z_2]$$

$$- e^{-\alpha_1 z_1} [4w_4 + 2\alpha_1 w_4 z_1] - e^{-\alpha_2 z_2} [4w_4 + 2\alpha_2 w_4 z_2] + 8w_4 \}$$

(B.18)

where

$$A = \frac{\alpha_1 \alpha_2}{8(K_0 + \theta)}$$

$$w_1 = 2\alpha_1 \alpha_2 (K_0 + \theta) + K_3 = 2w_4 + K_3$$

$$w_2 = \alpha_1 \alpha_2 (K_0 + \theta) + K_3 = w_4 + K_3$$

$$w_3 = \alpha_1 \alpha_2 (K_3 + \alpha_1 \alpha_2 K_0)$$

(B.19)

$$w_4 = \alpha_1 \alpha_2 (K_0 + \theta)$$

$$w_5 = \alpha_1 \alpha_2 (\alpha_1 \alpha_2 K_0 - K_3)$$

and similarly for the other parts;

ii) $\underline{z_1, z_2 < 0}$

$$I_5 = \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} e^{\alpha_1 x_1 + \alpha_2 x_2} [K_0 - \alpha_1 \theta x_1 - \alpha_2 \theta x_2 + (K_3 + \alpha_1 \alpha_2 K_0) x_1 x_2] dx_1 dx_2$$

$$F(z_1, z_2) = A \cdot I_5 = \frac{A}{\alpha_1^2 \alpha_2^2} e^{\alpha_1 z_1 + \alpha_2 z_2} [w_1 - \alpha_1 w_2 z_1 - \alpha_2 w_2 z_2 + w_3 z_1 z_2]$$

$$(B.20)$$

iii) $\underline{z_1 < 0, \ z_2 > 0}$

$$I_6 = \int_{-\infty}^{0} \int_{-\infty}^{z_1} e^{\alpha_1 x_1 + \alpha_2 x_2} [K_0 - \alpha_1 \theta x_1 - \alpha_2 \theta x_2 + (K_3 + \alpha_1 \alpha_2 K_0) x_1 x_2] dx_1 dx_2$$

$$= \frac{e^{\alpha_1 z_1}}{\alpha_1^2 \alpha_2^2} [w_1 - \alpha_1 w_2 z_1] \qquad (B.21)$$

$$I_7 = \int_{0}^{z_2} \int_{-\infty}^{z_1} e^{\alpha_1 x_1 - \alpha_2 x_2} [K_0 - \alpha_1 \theta x_1 + \alpha_2 \theta x_2 + (K_3 - \alpha_1 \alpha_2 K_0) x_1 x_2] dx_1 dx_2$$

$$= \frac{-e^{\alpha_1 z_1 - \alpha_2 z_2}}{\alpha_1^2 \alpha_2^2} \{2w_4 - K_3 - \alpha_1(w_4 - K_3)z_1 + (w_4 - K_3)\alpha_2 z_2 - w_5 z_1 z_2\}$$

$$+ \frac{e^{\alpha_1 z_1}}{\alpha_1^2 \alpha_2^2} [2w_4 - K_3 - \alpha_1(w_4 - K_3)z_1] \qquad (B.22)$$

$$F(z_1, z_2) = A(I_6 + I_7)$$

$$= \frac{A}{\alpha_1^2 \alpha_2^2} \{-e^{-\alpha_2 z_2 + \alpha_1 z_1}[2w_4 - K_3 - \alpha_1(w_4 - K_3)z_1 + \alpha_2(w_4 - K_3)z_2$$

$$- w_5 z_1 z_2] + e^{\alpha_1 z_1}[4w_4 - 2\alpha_1 w_4 z_1]\} \qquad (B.23)$$

SECOND-ORDER BIVARIATE SARGAN DENSITY

Let

$$f(x_1,x_2) = A \cdot e^{-\alpha_1|x_1|-\alpha_2|x_2|}[K_0+K_1|x_1|+K_2|x_2|+K_3x_1x_2+K_4|x_1x_2|$$

$$+ K_5x_1^2+K_6x_2^2] \qquad (C.1)$$

where

$$A = \frac{\alpha_1^3\alpha_2^3}{4B} \qquad (C.2)$$

$$B = \alpha_1^2\alpha_2^2K_0+\alpha_1\alpha_2^2K_1+\alpha_1^2\alpha_2K_2+\alpha_1\alpha_2K_4+2\alpha_2^2K_5+2\alpha_1^2K_6$$

Its marginal densitites are:

1) $\underline{x_1 > 0}$

$$L_1 = e^{-\alpha_1x_1}\int_0^\infty e^{-\alpha_2x_2}(K_0+K_1x_1+K_2x_2+(K_3+K_4)x_1x_2+K_5x_1^2+K_6x_2^2)dx_2$$

$$= e^{-\alpha_1 x_1}[\frac{K_0+K_1x_1+K_5x_1^2}{\alpha_2} + \frac{K_2+(K_3+K_4)x_1}{\alpha_2^2} + \frac{2x_6}{\alpha_3^2}] \qquad (C.3)$$

$$L_2 = e^{-\alpha_1 x_1} \int_{-\infty}^{0} e^{\alpha_2 x_2}(K_0+K_1x_1-K_2x_2+(K_3-K_4)x_1x_2+K_5x_1^2+K_6x_2^2)dx_2$$

$$= e^{-\alpha_1 x_1} [\frac{K_0+K_1x_1+K_5x_1^2}{\alpha_2} + \frac{K_2-(K_3-K_4)x_1}{\alpha_2^2} + \frac{2K_6}{\alpha_2^3}] \qquad (C.4)$$

ii)   and similarly for $x_1 < 0$, we get:

$$g_1(x_1) = \begin{cases} A(L_1+L_2) = \dfrac{2Ae^{-\alpha_1 x_1}}{\alpha_2^3} [\lambda_0+\lambda_1x_1+\lambda_2x_1^2] & \text{if } x_1 > 0 \\[4mm] = \dfrac{2Ae^{\alpha_1 x_1}}{\alpha_2^3} [\lambda_0-\lambda_1x_1+\lambda_2x_1^2] & \text{if } x_1 < 0 \end{cases}$$

$$(C.5)$$

where

$$\lambda_0 = \alpha_2^2 K_0+\alpha_2 K_2+2K_6$$

$$\lambda_1 = \alpha_2^2 K_1+\alpha_2 K_4$$

$$\lambda_2 = \alpha_2^2 K_5$$

for $g_1(X_1)$ to be second order Sargen requires:

$$\frac{\lambda_1}{\lambda_0} = \alpha_1\gamma_1 \qquad (C.6)$$

$$\frac{\lambda_2}{\lambda_0} = \alpha_1^2 \gamma_2 \tag{C.7}$$

substituting back into (C.5),

$$g_1(x_1) = \frac{\alpha_1 e^{-\alpha_1 |x_1|}}{2(1+\gamma_1+2\gamma_2)} (1+\alpha_1 \gamma_1 |x_1|+\alpha_1^2 \gamma_2 x_1^2) \tag{C.8}$$

and similarly for x_2:

$$g_2(x_2) = \frac{\alpha_2 e^{-\alpha_2 |x_2|}}{2(1+\xi_1+2\xi_2)} (1+\alpha_2 \xi_1 |x_2|+\alpha_2^2 \xi_2 x_2^2) \tag{C.9}$$

where

$$\xi_1 = \frac{\alpha_1^2 K_2 + \alpha_1 K_4}{\alpha_2(\alpha_1^2 K_0 + \alpha_1 K_1 + 2K_5)} \tag{C.10}$$

$$\xi_2 = \frac{\alpha_1^2 K_6}{\alpha_2^2(\alpha_1^2 K_0 + \alpha_1 K_1 + 2K_5)} \tag{C.11}$$

Equations (C.7) and (C.8) are second-order Sargan densities for X_1, and X_2 respectively. To ensure that first and second partial derivatives are continuous we set $\xi_1 = 1$, $\gamma_1 = 1$ (see Goldfeld and Quandt (1981)). Equations (C.6) and (C.10) result in:

$$\frac{\alpha_2^2 K_1 + \alpha_2 K_4}{\alpha_2^2 K_0 + \alpha_2 K_2 + 2K_6} = \alpha_1 \Rightarrow K_0 = \frac{K_1}{\alpha_1} - \frac{K_2}{\alpha_2} + \frac{K_4}{\alpha_1 \alpha_2} - \frac{2K_6}{\alpha_2^2} \qquad (C.12)$$

$$\frac{\alpha_1^2 K_2 + \alpha_1 K_4}{\alpha_1^2 K_0 + \alpha_1 K_1 + 2K_5} = \alpha_2 \Rightarrow K_0 = \frac{K_2}{\alpha_2} - \frac{K_1}{\alpha_1} + \frac{K_4}{\alpha_1 \alpha_2} - \frac{2K_5}{\alpha_1^2} \qquad (C.13)$$

That is;

$$\frac{K_1}{\alpha_1} - \frac{K_2}{\alpha_2} = \frac{K_5}{\alpha_1^2 \gamma_2} - \frac{K_6}{\alpha_2^2 \xi_2} \qquad (C.14)$$

Equation (C.11) can be rewritten as:

$$K_0 = \frac{K_6}{\alpha_2^2 \xi_2} - \frac{K_1}{\alpha_1} - \frac{2K_5}{\alpha_1^2} \qquad (C.15)$$

and from (C.7) we get:

$$\frac{\alpha_2 K_5}{\alpha_2^2 K_0 + \alpha_2 K_2 + 2K_6} = \alpha_1^2 \gamma_2 \Rightarrow \frac{K_5}{\alpha_1^2 \gamma_2} - \frac{K_2}{\alpha_2} - \frac{2K_6}{\alpha_2^2} \qquad (C.16)$$

Using equations (C.14) - (C.16), we get:

$$\frac{K_5}{\alpha_1^2} (1 + \frac{1}{\gamma_2}) = \frac{K_6}{\alpha_2^2} (1 + \frac{1}{\xi_2}) = \delta \text{ (say)} \qquad (C.17)$$

Then;

$$K_5 = \frac{\alpha_1^2 \gamma_2 \delta}{1 + \gamma_2} \quad , \quad K_6 = \frac{\alpha_2^2 \xi_2 \delta}{1 + \xi_2} \qquad (C.18)$$

and,

$$K_4 = \alpha_1 \alpha_2 K_0 + \alpha_1 \alpha_2 \delta \left(\frac{\gamma_2}{1+\gamma_2} + \frac{\xi_2}{1+\xi_2} \right) = \alpha_1 \alpha_2 v_3 \qquad (C.19)$$

$$v_4 = \frac{\gamma_2}{1+\gamma_2} \quad , \quad v_5 = \frac{\xi_2}{1+\xi_2} \qquad (C.20)$$

$$v_3 = K_0 + \delta (v_4 + v_5) \qquad (C.21)$$

Therefore

$$K_1 = \alpha_1 v_1 \quad , \quad K_2 = \alpha_2 v_2$$

$$v_1 = \delta \left(\frac{1}{1+\xi_2} - \frac{2\gamma_2}{1+\gamma_2} \right) - K_0 \qquad (C.22)$$

$$v_2 = \delta \left(\frac{1}{1+\gamma_2} - \frac{2\xi_2}{1+\xi_2} \right) - K_0 \qquad (C.23)$$

Using the above relationships, the joint density can be rewritten as

$$f(x_1,x_2) = \frac{\alpha_1 \alpha_2}{8\delta} e^{-\alpha_1 |x_1| - \alpha_2 |x_2|} [K_0 + \alpha_1 v_1 |x_1| + \alpha_2 v_2 |x_2| + K_3 x_1 x_2 + \alpha_1 \alpha_2 v_3 |x_1 x_2|$$

$$+ \alpha_1^2 \delta v_4 x_1^2 + \alpha_2^2 \delta v_5 x_2^2] \qquad (C.24)$$

The covariance of X_1, X_2 will therefore be derived as follows:

$$M_1 = \int_{-\infty}^{0} x_2 e^{\alpha_2 x_2} \int_{-\infty}^{0} x_1 e^{\alpha_1 x_1} [K_0 - \alpha_1 v_1 x_1 - \alpha_2 v_2 x_2 + (K_3 + \alpha_1 \alpha_2 v_3) x_1 x_2 + \alpha_1^2 \delta v_4 x_1^2$$

$$+ \alpha_2^2 v_5 \delta x_2^2] dx_1 dx_2$$

$$= \int_{-\infty}^{0} x_2 e^{\alpha_2 x_2} \left\{ - \frac{K_0 - \alpha_2 v_2 x_2 + \alpha_2^2 v_5 \delta x_2^2}{\alpha_1^2} + \frac{2[-\alpha_1 v_1 + (K_3 + \alpha_1 \alpha_2 v_3) x_2]}{\alpha_1^3} \right.$$

$$\left. - \frac{6\alpha_1^2 \delta v_4}{\alpha_1^4} \right\} \ dx_2$$

$$= \frac{\alpha_1^2 K_0 + 2\alpha_1^2 v_1 + 6\alpha_1^2 \delta v_4}{\alpha_1^4 \alpha_2^2} + 2 \ \frac{\alpha_1 \alpha_2 v_2 + 2(K_3 + \alpha_1 \alpha_2 v_3)}{\alpha_1^3 \alpha_2^3} + \frac{6\alpha_2^2 \delta v_5}{\alpha_2^4 \alpha_1^2} \qquad (C.25)$$

and similarly for the other parts, i.e.:

$$M_2 = \frac{-(\alpha_1^2 K_0 + 2\alpha_1^2 v_1 + 6\alpha_1^2 \delta v_4)}{\alpha_1^4 \alpha_2^2} + \frac{2[-\alpha_1 \alpha_2 v_2 + 2(K_3 - \alpha_1 \alpha_2 v_3)]}{\alpha_1^3 \alpha_2^3} - \frac{6\alpha_2^2 \delta v_5}{\alpha_1^2 \alpha_2^4}$$

$$(C.26)$$

$$M_3 = \frac{-(\alpha_1^2 K_0 + 2\alpha_1^2 v_1 + 6\alpha_1^2 \delta v_4)}{\alpha_1^4 \alpha_2^2} + \frac{2[-\alpha_1 \alpha_2 v_2 + 2(K_3 - \alpha_1 \alpha_2 v_3)]}{\alpha_1^3 \alpha_2^3} - \frac{6\alpha_2^2 \delta v_5}{\alpha_1^2 \alpha_2^4}$$

$$(C.27)$$

$$M_4 = \frac{\alpha_1^2 K_0 + 2\alpha_1^2 v_1 + 6\alpha_1^2 \delta v_4}{\alpha_1^4 \alpha_2^2} + \frac{2[\alpha_1 \alpha_2 v_2 + 2(K_3 + \alpha_1 \alpha_2 v_3)]}{\alpha_1^3 \alpha_2^3} + \frac{6\alpha_2^2 \delta v_5}{\alpha_1^2 \alpha_2^4}$$

$$(C.28)$$

$$COV(X_1, X_2) = \frac{\alpha_1 \alpha_2}{8\delta} (M_1 + M_2 + M_3 + M_4) = \frac{2K_3}{\delta \alpha_1^2 \alpha_2^2} \qquad (C.29)$$

and the correlation coefficient is:

$$\rho = \frac{2K_3 / \alpha_1^2 \alpha_2^2 \delta}{\sqrt{\dfrac{16(1 + 3\gamma_2)(1 + 3\xi_2)}{\alpha_1^2 \alpha_2^2 (1 + \gamma_2)(1 + \xi_2)}}} = \frac{K_3}{2\alpha_1 \alpha_2 \delta} \sqrt{\frac{(1 + \gamma_2)(1 + \xi_2)}{(1 + 3\gamma_2)(1 + 3\xi_2)}}$$

$$(3.30)$$

Chapter 6

Conclusions

The main objective of this dissertation was to investigate the adequacy of the Sargan distribution as an approximation to the normal in econometric models. The normal distribution is widely assumed, in part because it often leads to simple results. However, in some models the normal distribution is computationally more complicated than the alternative distributions, such as the Sargan distribution, whose c.d.f. can be expressed in closed form. Thus there may be a computational benefit to using an approximation to the normal in certain models, and we ask what the cost might be.

In chapters 3 and 4 we showed that the univariate Sargan distribution provides a very close approximation to the normal, in the sense that their densities are quite close to each other. This is especially so if a second or higher-order Sargan distribution is used. Such comparisons have also previously been made by Goldfeld and Quandt (1981) and by Missiakoulis (1983).

However, the fact that the densities are close does not necessarily imply that MLE's based on the Sargan distribution will have properties similar to the MLE's based on the normal distribution. Therefore, we have considered a variety of models, and asked what the

cost is of assuming the errors to have a Sargan distribution, if in fact they have a normal distribution. Our definition of cost is the asymptotic bias (inconsistency) of the resulting estimates.

This cost is model dependent; in other words, it is different in different models. There is no cost in the linear regression model. In fact, our result for the linear regression model is stronger: the Sargan MLE's are consistent regardless of the true distribution of the error terms, so long as it is symmetric. However, in the so-called sample selection models, in which the c.d.f. of the errors appears in the likelihood function, there is a cost to an incorrect assumption of the distribution of the errors. How high the cost is depends upon how complete the sample is; the cost is higher when the sample is more highly censored or truncated. For example, in the censored regression model the asymptotic bias varies positively with the degree of censoring. As the sample becomes more complete, the bias disappears, which is consistent with our result for the fully observed sample (the linear regression model).

The same kind of results occur for the truncated dependent variable model, except that the bias is generally larger in the truncated model than in the censored model. This is consistent with the reasonable intuition that additional information helps in reducing the bias of the estimates.

Another result which is consistent with this intuition is that the bias of the estimates is virtually always smaller when the error variance is known than it is when the error variance is unknown. The bias is sometimes large enough to be a serious problem, but it is

relatively minor for samples that are 50% complete in the censored case, or 75% complete in the truncated case.

The same kinds of results are true for higher-order Sargan distributions, but the bias is smaller when a higher-order Sargan distribution is assumed. Therefore, as should be expected, it is much safer to approximate the normal distribution by the second-order Sargan distribution than by the first-order Sargan.

Another interesting result is that it is much less costly (in terms of asymptotic bias) to approximate the normal distribution with the Sargan that vice-versa. In the linear regression model, there is no bias either way. However, in our more complicated models, the bias caused by assuming Sargan when normal is true is much smaller than the bias caused by assuming normal when Sargan is true. It is not apparent why this should be the case.

The overall conclusion from our study of the univariate Sargan distribution is that one should not use the Sargan distribution, if one really believes that the normal distribution is correct. Any computational savings are not worth the cost, in terms of asymptotic bias and resulting incorrect inferences.

On the other hand, while models depending on the univariate normal c.d.f. (e.g., the Tobit model) are not terribly complicated computationally, models that involve the multivariate normal c.d.f. (e.g., a multi-market disequilibrium model) are still very difficult to estimate. Thus a multivariate Sargan distribution might be more valuable, in terms of potential computational savings, than the univariate Sargan distribution. We have considered a bivariate

distribution, whose marginals are univariate Sargan, and defined it to be a bivariate Sargan distribution. We have shown that our constructed bivariate Sargan density is very close to the bivariate standard normal density except around the turning points, although it is not very far off at these intervals either.

Also we proved in the seemingly unrelated regressions model that the estimates based on the bivariate Sargan density are asymptotically unbiased (consistent) regardless of the true distribution of the error terms, as long as it is symmetric. Presumably this is not so in the more complex models. Although we did not investigate such models in detail, this could be done by a straight forward extension of the methods used earlier in the thesis.

The main pronlem with our bivariate Sargan density is the limited possible range of the correlation coefficient, ρ. It is not very interesting to limit attention to bivariate or multivariate Sargan distributions that exhibit almost no correlation, but this is necessary to keep the density non-negative over its entire range. An interesting unanswered question is whether ignoring this restriction would cause problems in actual empirical work. For example, it would certainly not matter in any practical sense if the Sargan density were negative, ten or twenty standard deviations from the mean.

BIBLIOGRAPHY

Arabmazar, A. and P. Schmidt (1982),"An Investigation of the Robustness of the Tobit Estimator to Non-Normality," *Econometrica*, 50, 1055-1063.

Goldberger, A. S. (1980),"Abnormal Selection Bias" SSRI Discussion Paper 8006, University of Wisconsin, Madison.

Goldfeld, S. M. and R. E. Quandt (1981),"Econometric Modeling with Non-Normal Disturbances," *Journal of Econometrics*, 17, 141-155.

Grobner W. and N. Hofreiter (1958), " *Integraltafel, Zweiter Teil Bestimate Integrale*," Springer-Verlag in Vienna, Austria.

Hechman, J. (1976), "The Common Structure of Statistaical Models of Truncation, Sample Selection and Limited Dependent Variables, aand a Simple Estimators for Such Models," *Annals of Economics and Social Measurement*, 5, 475-492.

Johnson, N.L. and S. Kotz (1970), Continuous Univariate Distributions, Vol 1., New York, Wiley.

Missiakoulis, S. (1983),"Sargan Densities: Which One?," *Journal of Econometrics*, 23, 223-234.

Stone, M. H. (1962),"A Generalized Weierstrass Approximation Theorem," in *Studies in Modern Analysis*, Vol I, ed. R. Greighton Buck, Englewood Cliffs, New Jersey: Prentice Hall.