

Figure 9.1 Estimates of α_1 as a function of S

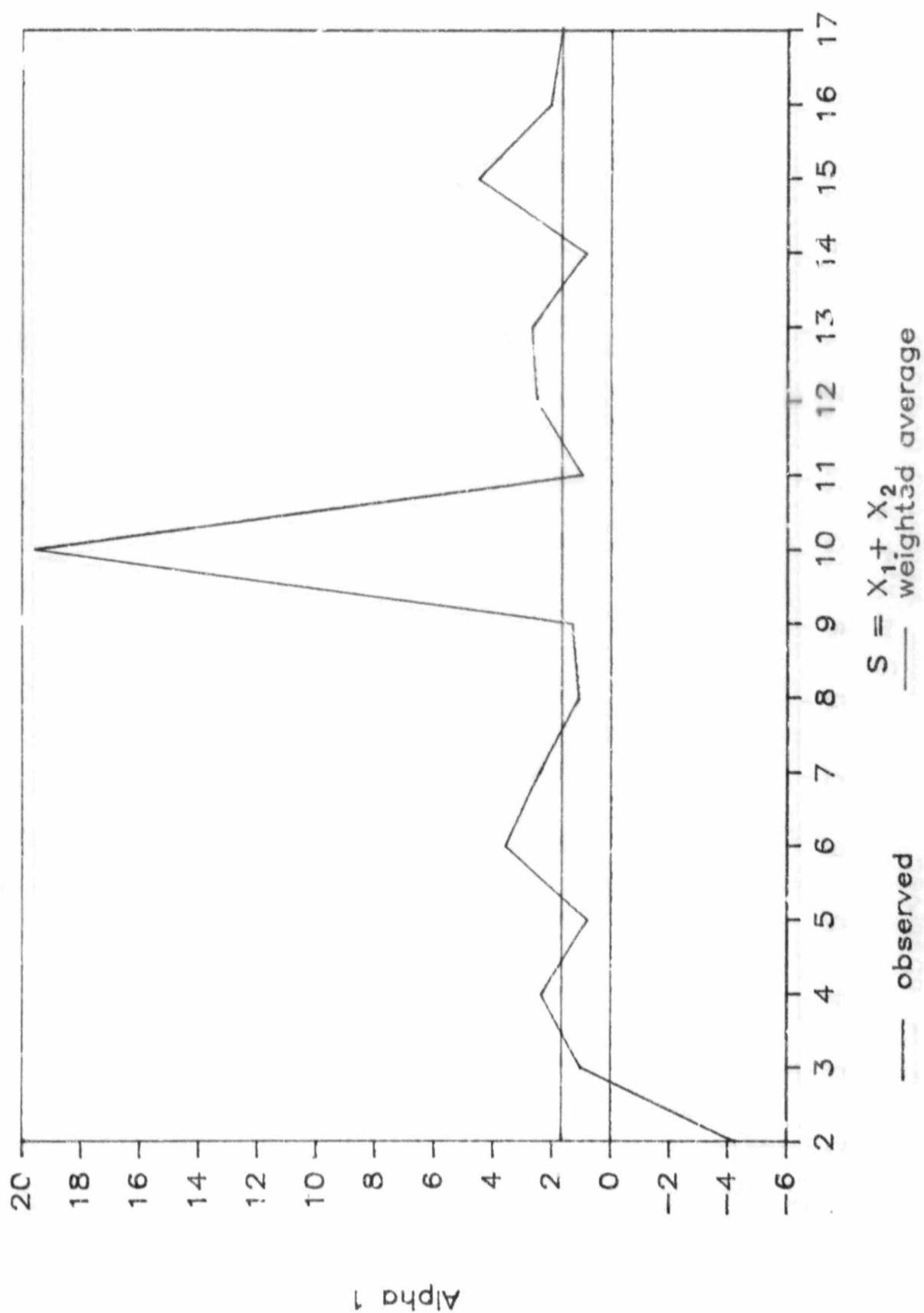
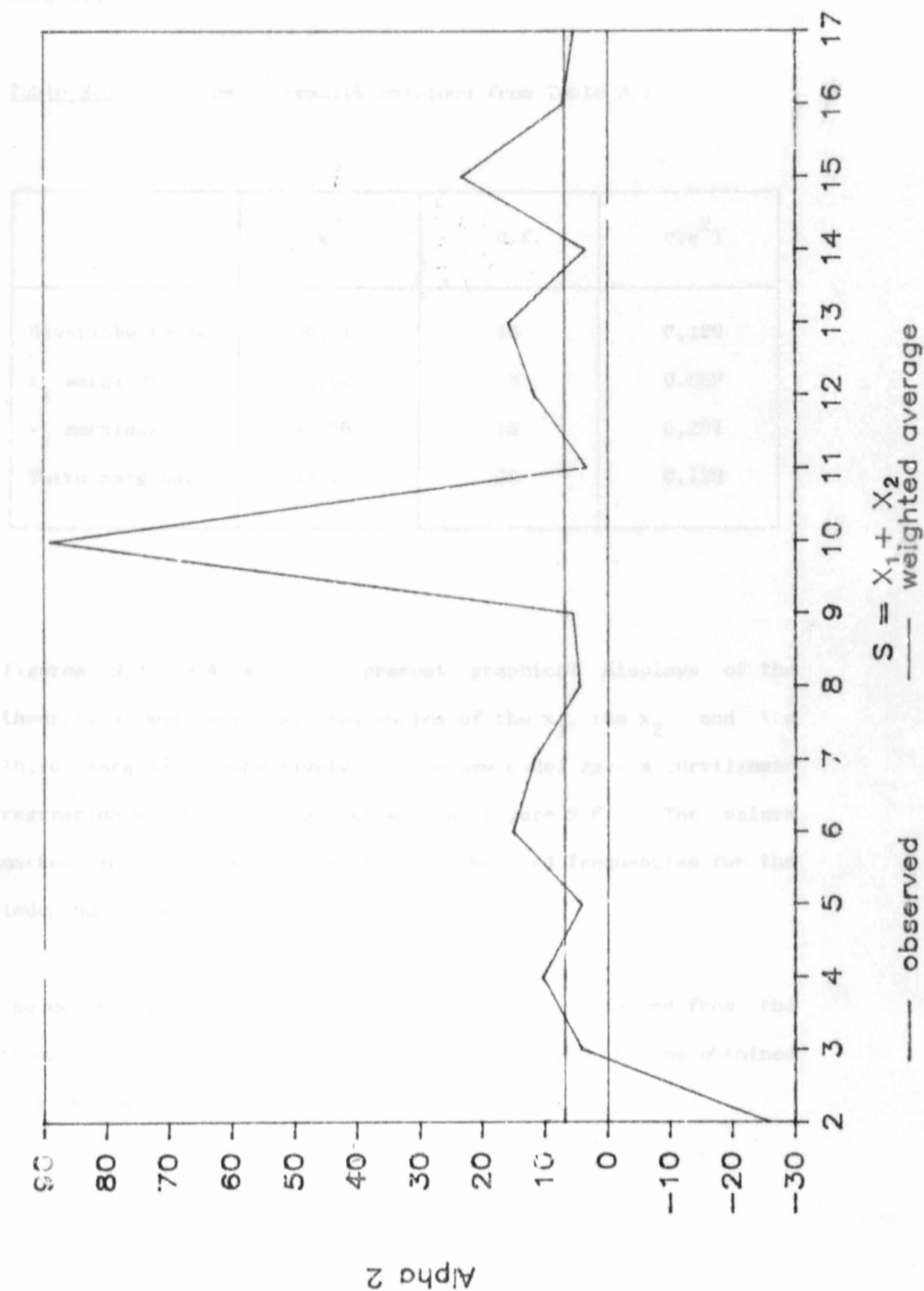


Figure 9.2 Estimates of α_2 as a function of S



The results presented in Table 9.3 were obtained from the analysis of the whole table, the marginals x_1 and x_2 and the third marginal.

Table 9.3 The χ^2 results obtained from Table 9.1

	χ^2	d.f.	$P(\chi^2)$
Bivariate table	90,71	76	0,120
x_1 marginal	2,50	6	0,869
x_2 marginal	22,56	19	0,257
Third marginal	33,1	25	0,129

Figures 9.3, 9.4 and 9.5 present graphical displays of the theoretical and empirical frequencies of the x_1 , the x_2 and the third marginal respectively. The new model gave a curvilinear regression of $E(x_2|x_1)$, as shown in Figure 9.6. The values marked in brackets represent the observed frequencies for the independent variable.

The correlation coefficient between x_1 and x_2 , obtained from the theoretical distribution, is $\rho = 0,3534$ and the one obtained from the empirical data is 0,4352.

Figure 9.3 The x_1 marginal, data of Bates and Neyman (1952)

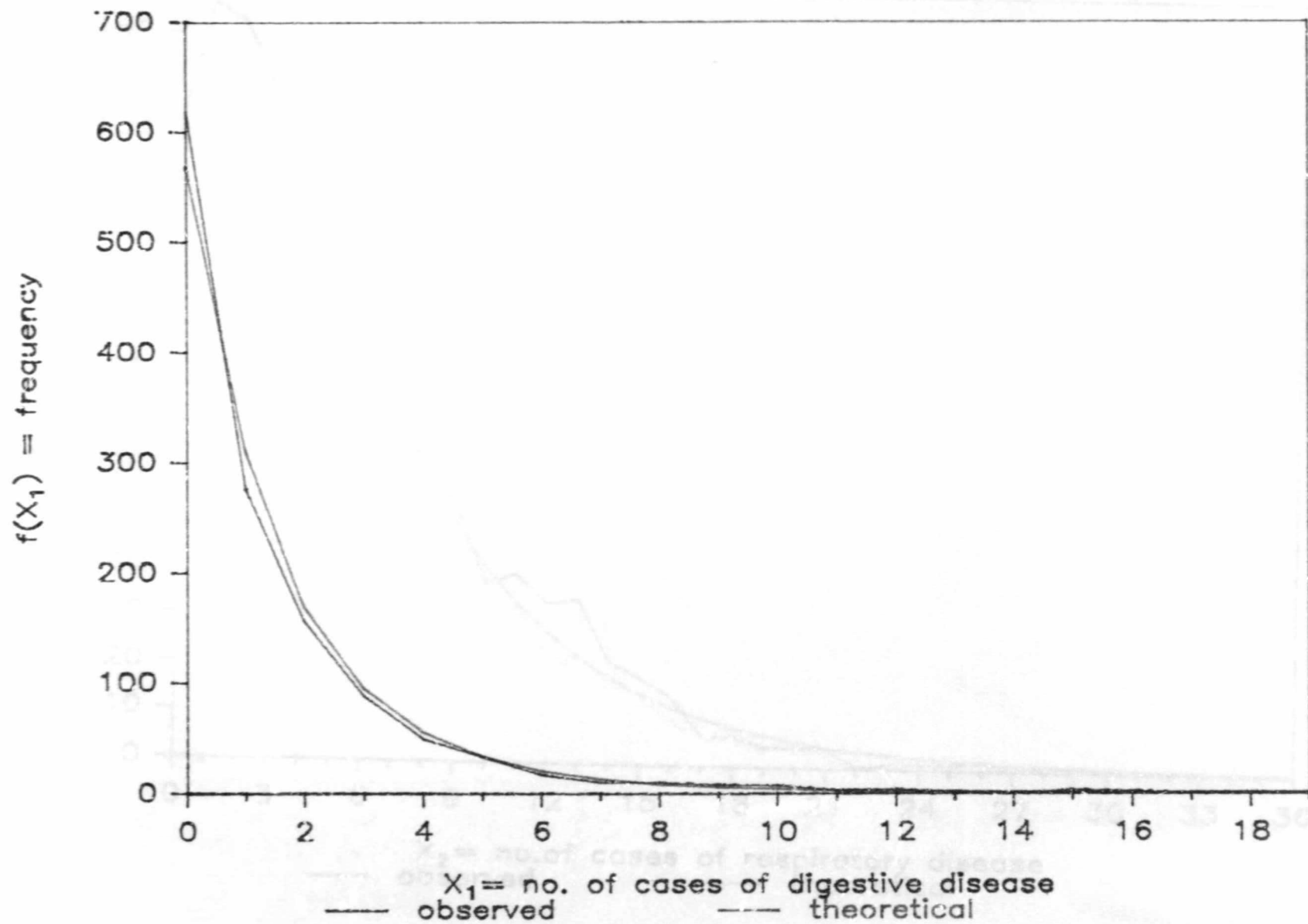


Figure 9.4 The x_2 marginal, data of Bates and Neyman (1952)

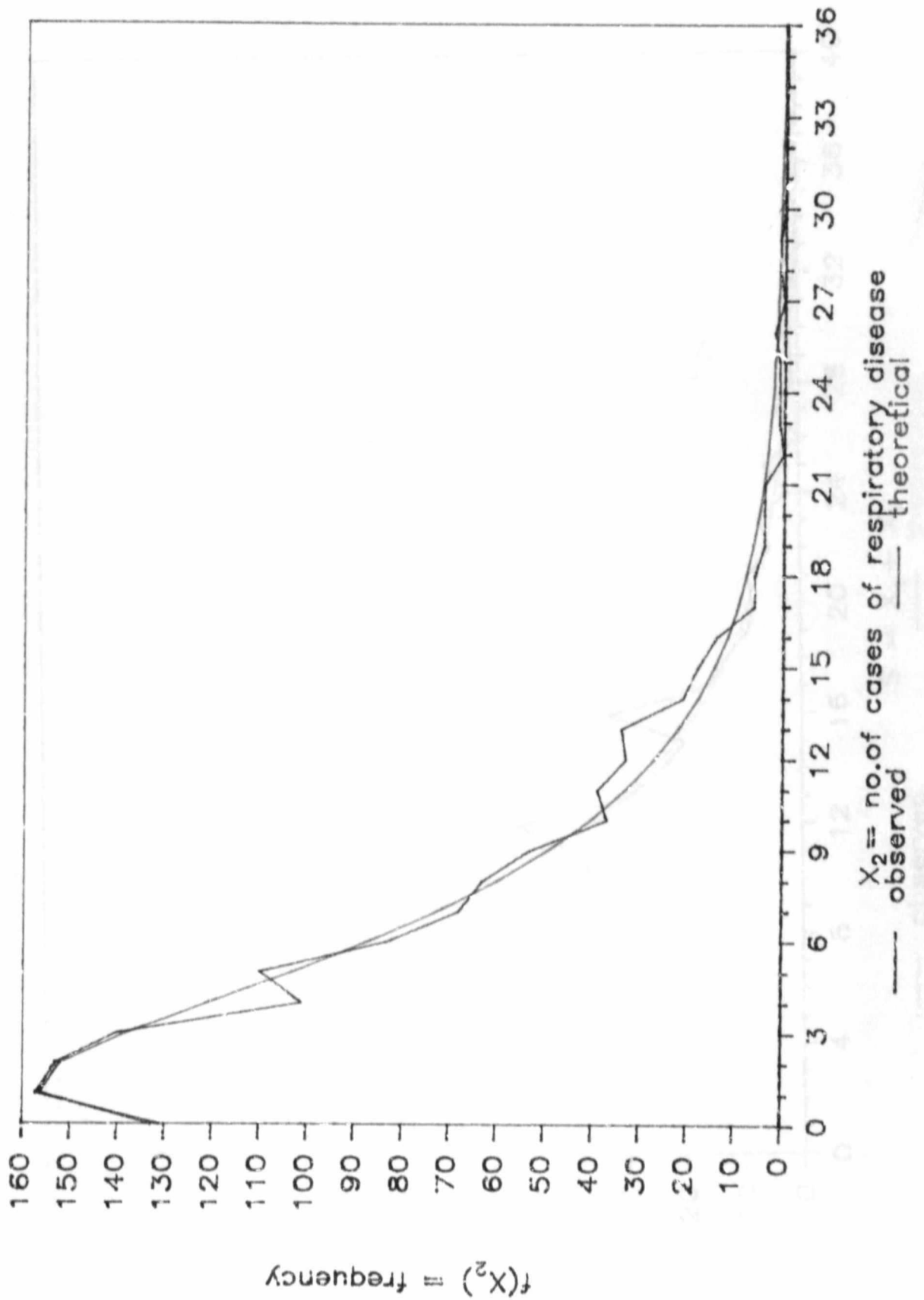


Figure 9.5 The third marginal, data of Bates and Neyman (1952)

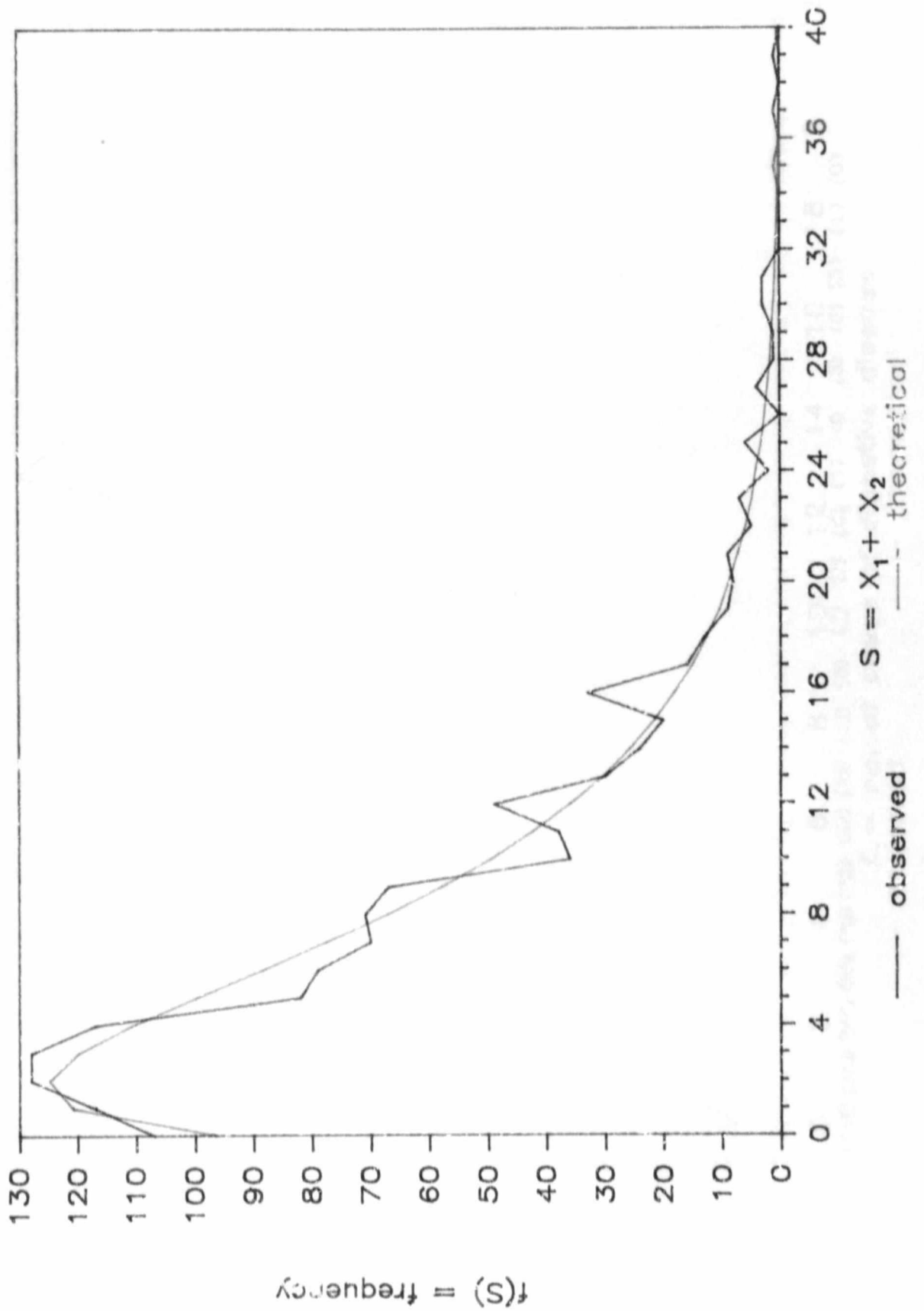
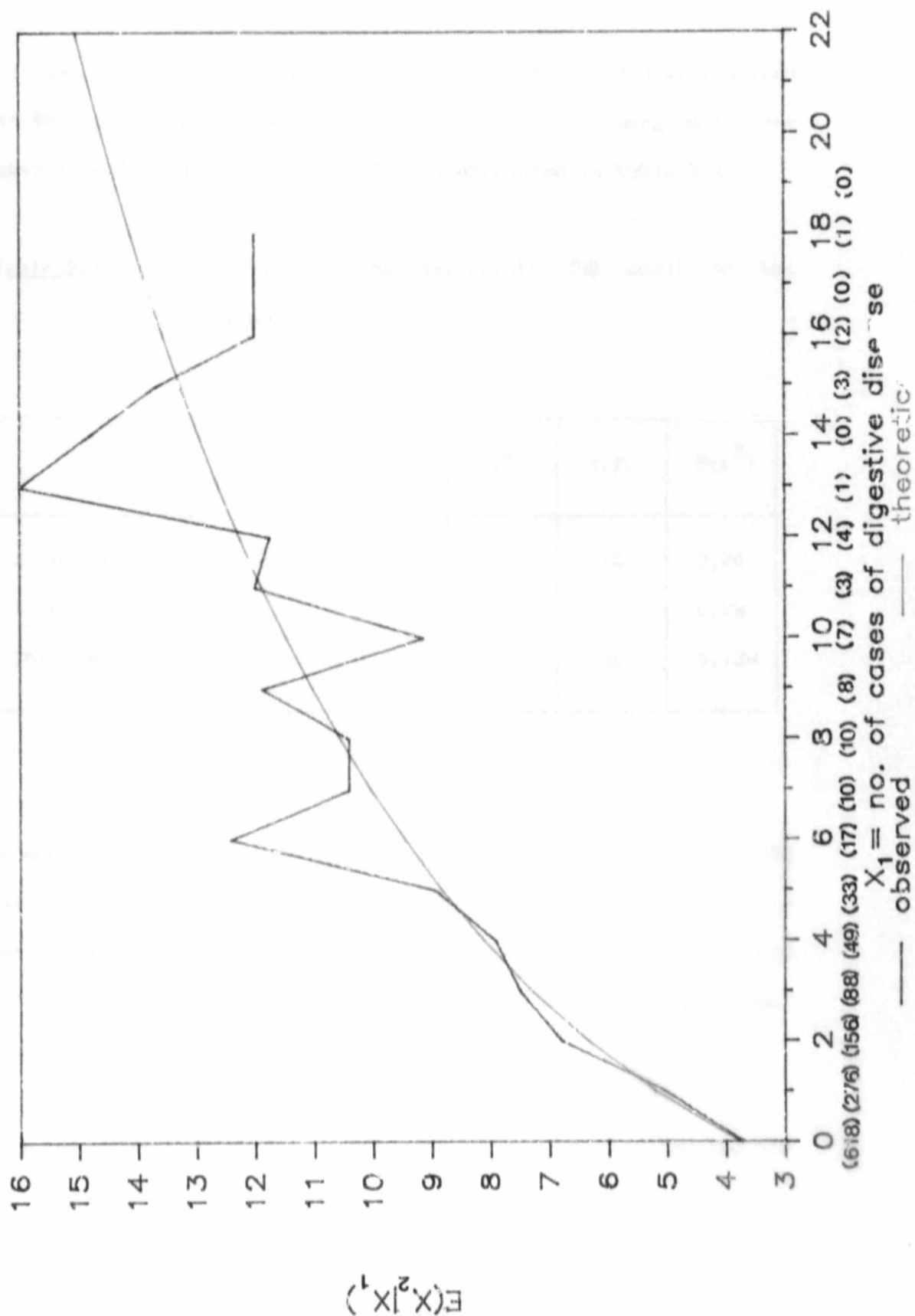


Figure 9.6 The regression curve $E(x_2|x_1)$,
data of Bates and Neyman (1952)



These results are by far superior to the results obtained by Bates and Neyman (1952) and Mitchell and Paulson (1981), as previously stated.

It is possible to fit univariate Negative Binomial distributions to the third marginal and to the x_1 and x_2 marginals. The results obtained in terms of χ^2 are presented in Table 9.4.

Table 9.4 The fit of the univariate NBD model to the marginals

	$\hat{\gamma}$	$\hat{\theta}$	χ^2	d.f.	$P(\chi^2)$
x_1 marginal	0,53	0,724	12,70	10	0,24
x_2 marginal	1,69	0,759	29,47	26	0,29
Third marginal	1,54	0,8136	33,1	25	0,129

As mentioned by Chatfield, Ehrenberg and Goodhardt (1966) the NBD model was successfully used in a variety of practical fields as "in the study of accident statistics, in certain sociological birth and death and contagious processes and in some operational research theory."

From the univariate empirical data one can successfully fit the NBD

model to the marginals x_1 and x_2 and the third marginal. However from the theoretical development of the bivariate distribution it is obvious that the above marginals do not follow an NBD distribution.

By comparing Tables 9.3 and 9.4 one can see that for the x_1 marginal we obtain a better fit from the bivariate table and for the other marginals we obtain a fit of the same order.

In Table 9.3 the variable x_1 represents the number of children of children for the first period 2-7 years, x_2 represents the number of children of children for the second period 8-11 years and $x_3 = x_1 + x_2$.

The new model was fitted to the data set by fitting the NBD for the third marginal and the bivariate table. The results are shown in Table 9.4. The theoretical distribution was compared with the observed data. The theoretical distribution was compared with the observed data in Table 9.4.

9.2 The Data Set from the Children Accident Field,

used by Mellinger et al. (1965)

Mellinger et al. (1965) applied the Bates-Neyman model (the bivariate NBD) to four sets of accident data obtained in a study of childhood accidents. The study was performed on a few groups of children whose accident records were taken over two consecutive four-year periods.

The data set of Table 9.5 contains information on 621 children with accident histories over the periods of 4 to 7 years and 8 to 11 years. Mellinger et al. (1965) fit the Bates-Neyman (1952) model to these observations with the estimates $\hat{\alpha} = 3,18$, $\hat{\beta} = 2,81$ and $\hat{A} = 1,18$ and obtained $\chi^2 = 16,15$ with d.f. = 24 and $P(\chi^2) = 0,88$.

In Table 9.5 the variable x_1 represents the number of injuries of children for the first period 4-7 years, x_2 represents the number of injuries of children for the second period 8-11 years and $S = x_1 + x_2$.

The new model was applied to this data set by fitting the NBD for the third marginal with the parameters $\hat{\gamma} = 3,587$ and $\hat{\theta} = 0,406$; and the BBD model for the ascending diagonal arrays with the parameters $\hat{\alpha}_1 = 12,27$ and $\hat{\alpha}_2 = 11,48$. The theoretical frequencies are presented together with the empirical ones in Table 9.5.

Table 9.5

The data set of Melinger et al. (1965) with

$$\hat{\gamma} = 3,587, \hat{\theta} = 0,406, \hat{\alpha}_1 = 12,27 \text{ and } \hat{\alpha}_2 = 11,46.$$

s	f(s)	$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	$f(x_1)$
0	101 95,87	0	101 95,87	76 72,13	35 36,01	15 15,09	7 5,76	3 2,08	3 0,73	0,36	240 228,03
1	143 139,61	1	67 67,48	61 62,31	32 36,42	14 17,32	12 7,34	4 2,90	1 1,09	1 0,60	192 195,46
2	120 130,01	2	24 31,69	36 34,25	22 22,72	15 12,00	6 5,55	1 2,36	2 0,95	1 0,58	107 110,11
3	93 98,30	3	10 12,54	19 15,38	10 11,33	5 6,54	2 3,26	4 1,49	0 0,63	2 0,43	52 51,60
4	63 65,72	4	1 4,54	7 6,18	3 4,97	4 3,10	2 1,55	0 0,80	0 0,36	0 0,25	17 21,85
5	49 40,49	5	2 1,56	1 2,32	4 2,61	2 1,34	0 0,76	0 0,38	0 0,18	0 0,15	9 8,70
6	52 51,00	6	1 0,77	1 1,27	1 1,22	1 0,87	0 0,55	0 0,27	0 0,15	0 0,14	4 5,26
		$f(x_2)$	206 214,45	201 193,84	107 114,68	56 56,26	29 24,87	12 10,30	6 4,09	4 2,51	621 621,00

The results presented in Table 9.6 were obtained from the analysis of the whole table, the marginals x_1 and x_2 and the third marginal.

Table 9.6 The χ^2 results obtained from Table 9.5

	χ^2	d.f.	$P(\chi^2)$
Bivariate table	18,27	23	0,743
x_1 marginal	2,17	2	0,338
x_2 marginal	2,17	2	0,338
Third marginal	7,44	7	0,385

From the graphical display of the x_1 marginal, the x_2 marginal and the third marginal given in Figures 9.7, 9.8 and 9.9 respectively, one can see that the above model gives a reasonably good fit to the data.

The regression $E(x_2|x_1)$ is presented in Figure 9.10. The differences between the observed and the theoretical curves are numerically small, but are blown up due to the scale chosen. The main purpose was to show non-linearity. When inspecting this figure one should bear in mind that regression points corresponding to large values of the independent variable depend on a low number of observations.

The correlation coefficient between x_1 and x_2 obtained from the theoretical distribution is $\rho = 0,198$ and that from the empirical data is 0,185.

Figure 9.7 The x_1 marginal, data of Mellinger et al. (1965)

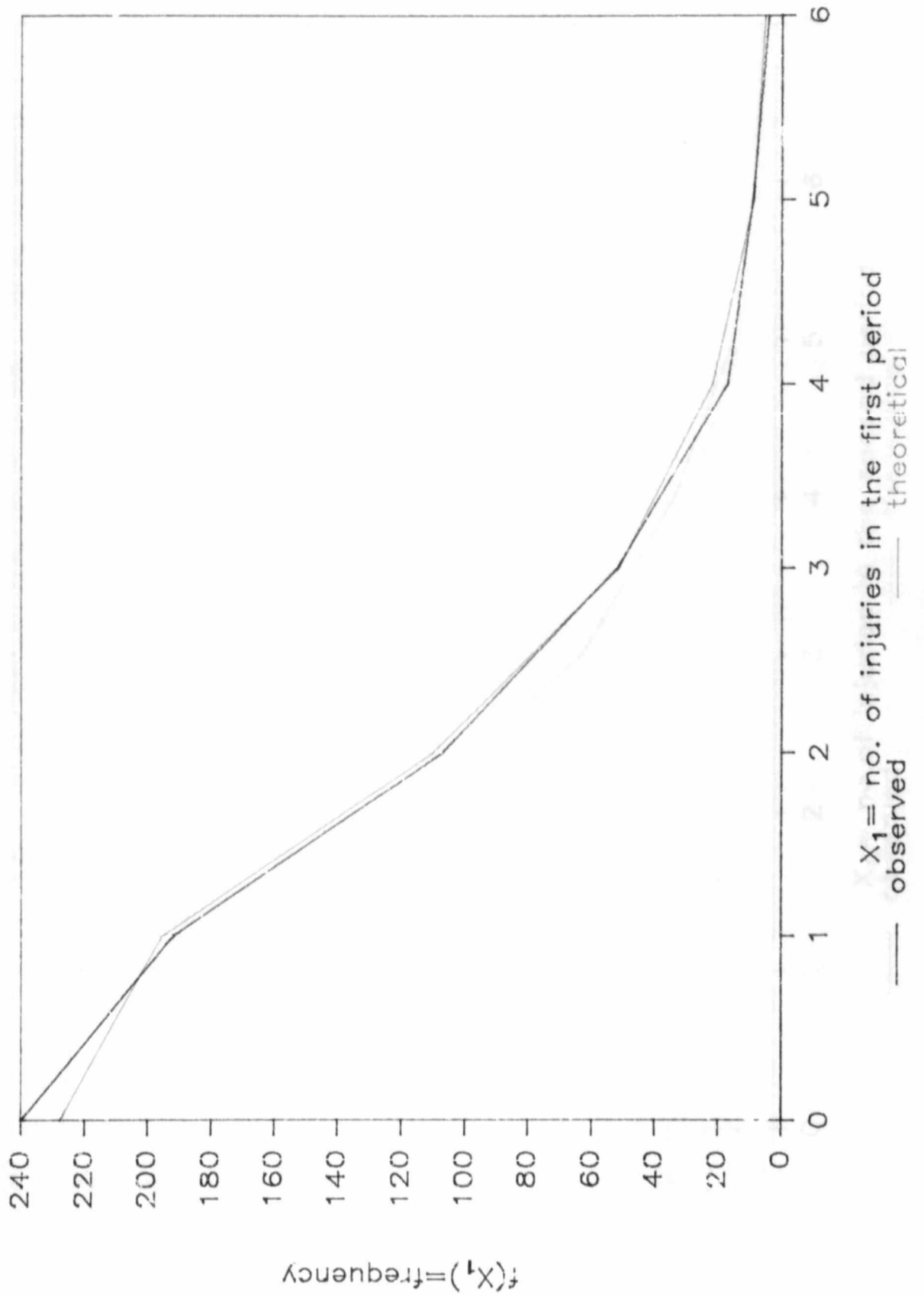


Figure 9.8

The x_2 marginal, data of Mellinger et al. (1965)

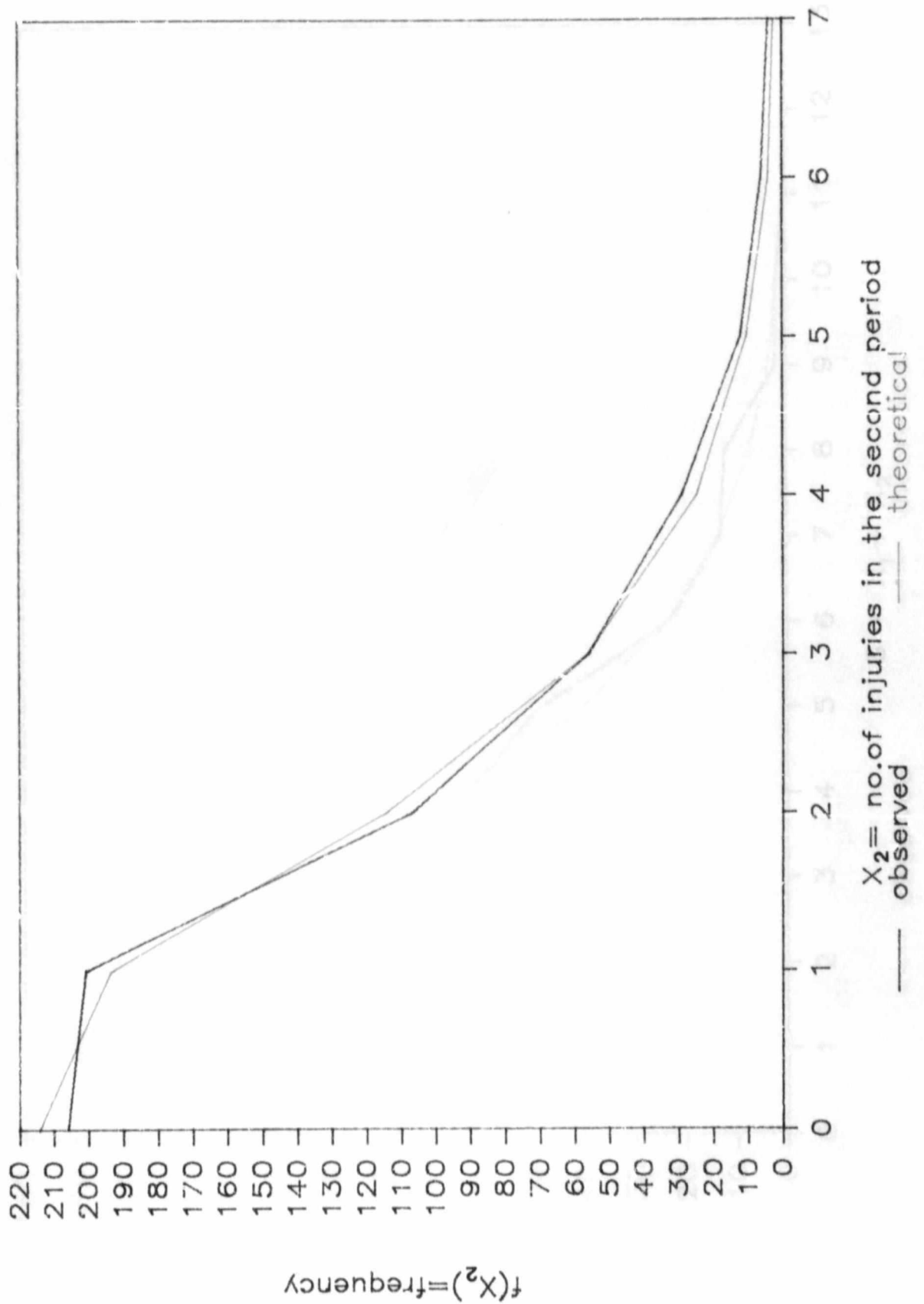


Figure 9.9 The third marginal, data of Mellinger et al. (1965)

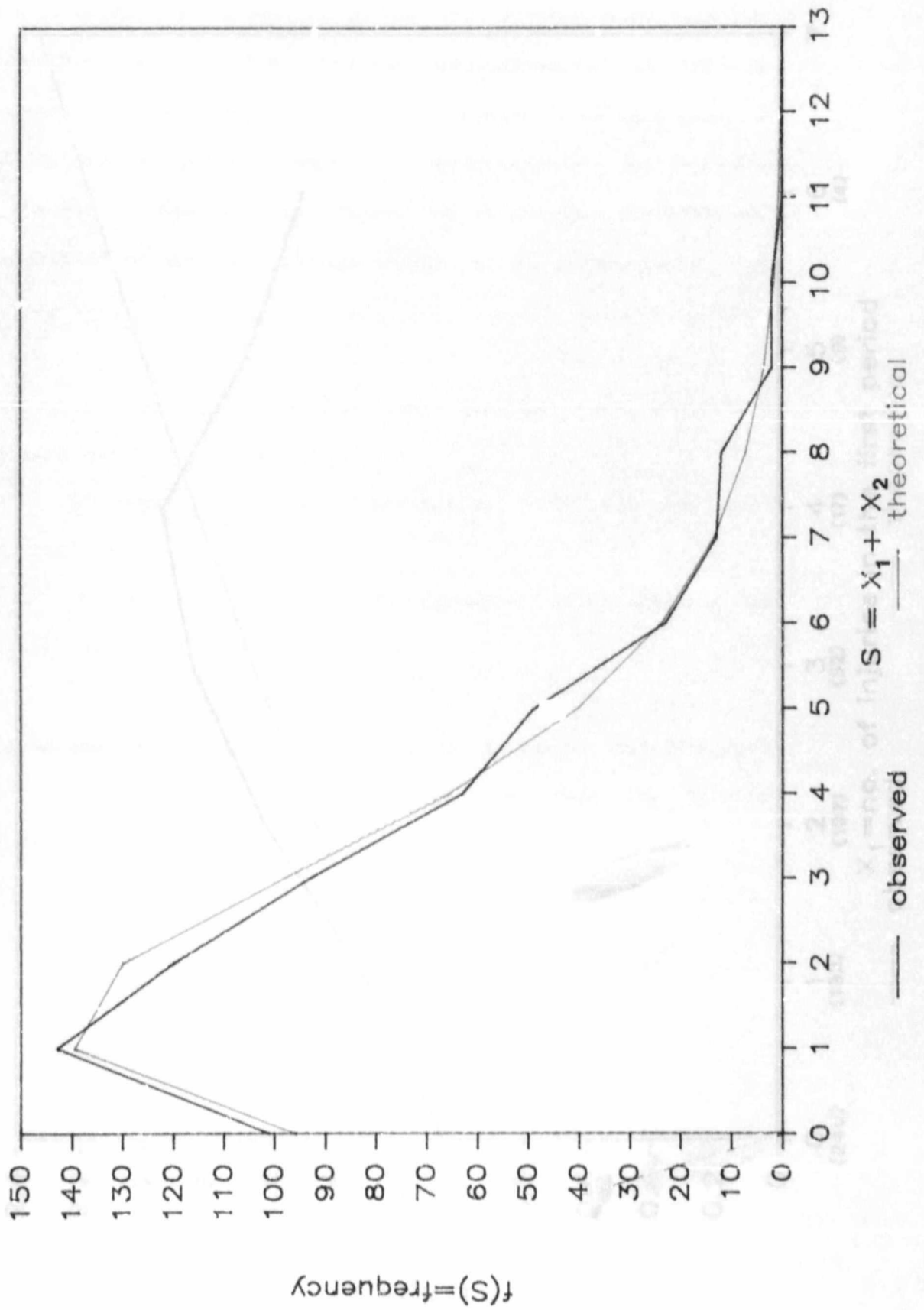
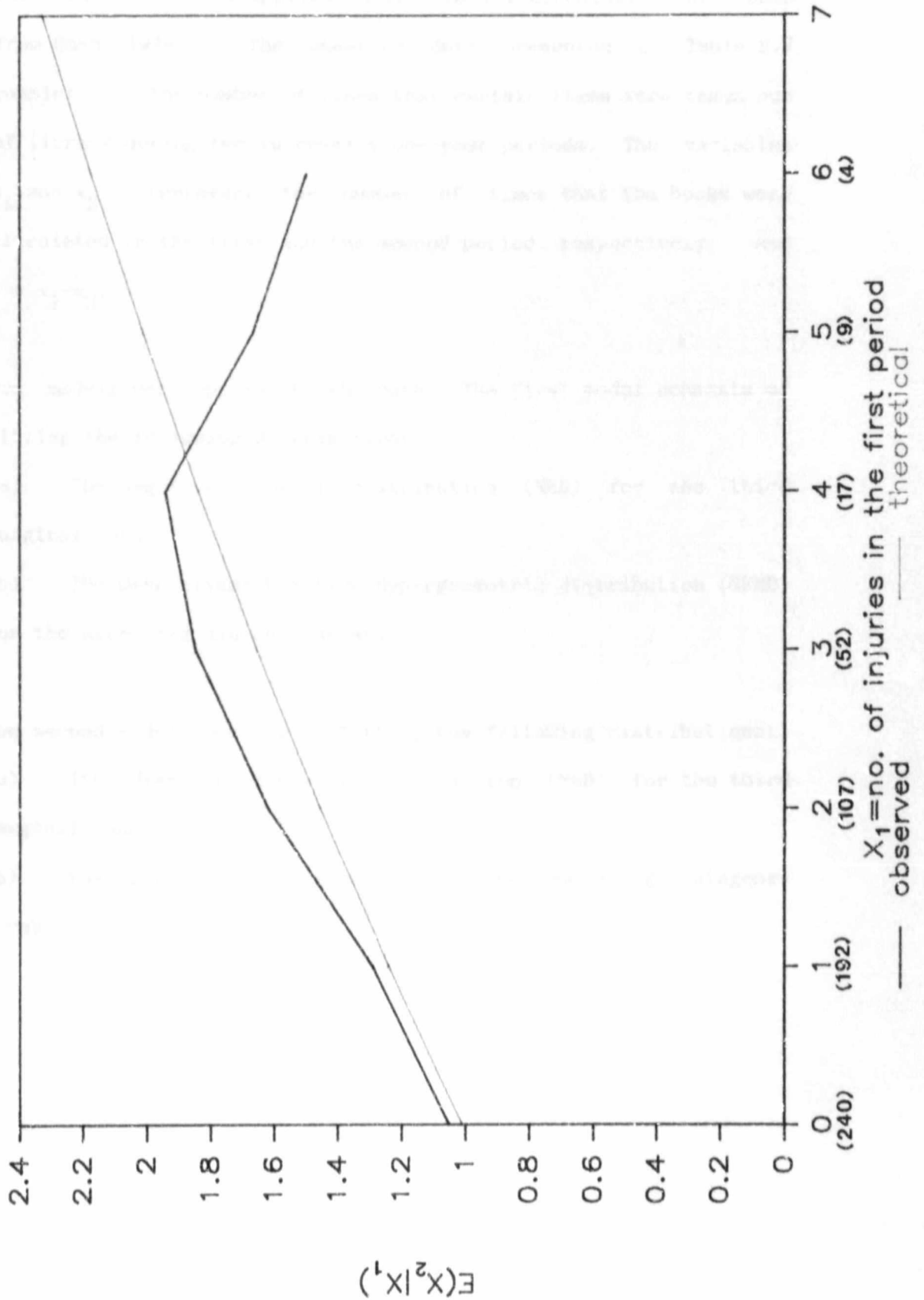


Figure 9.10 The regression curve $E(x_2|x_1)$, Mellinger et al. (1965)



9.3 The Data Set from the Library Circulation Field,
used by Chen (1976)

The new theory is applied on the library circulation data taken from Chen (1976). The observed data presented in Table 9.7 consist of the number of times that certain items were taken out of library during two successive one-year periods. The variables x_1 and x_2 represent the number of times that the books were circulated in the first and the second period respectively and $S = x_1 + x_2$.

Two models were applied to the data. The first model consists of fitting the following distributions:

- (a) The Negative Binomial distribution (NBD) for the third marginal and
- (b) The Generalized Positive Hypergeometric distribution (GPHD) for the ascending diagonal arrays.

The second model consists of fitting the following distributions:

- (a) The Negative Binomial distribution (NBD) for the third marginal and
- (b) The Binomial distribution for the ascending diagonal arrays.

Table 9.7 The data set of Chen (1976), the NBD - GPHD model

with $\hat{\gamma} = 3,91$, $\hat{\theta} = 0,56$, $\hat{\alpha}_1 = -39,32$ and $\hat{\alpha}_2 = -38,867$.

S	f(S)	$x_2 \backslash x_1$	0	1	2	3	4	5	6	7	8	9	≥ 10	f(x ₂)
0	23 25,22	0	23 25,22	28 27,77	17 18,96	9 10,24	6 4,79	1 2,02	1 0,79	0 0,29	0 0,10	1 0,03	0 0,01	86 90,22
1	64 55,22	1	36 27,45	36 38,45	38 32,01	11 20,49	10 11,11	7 5,35	2 2,35	0 0,96	0 0,37	0 0,13	0 0,08	140 138,75
2	76 75,93	2	23 18,52	31 31,63	28 31,19	27 23,16	15 14,34	11 7,79	2 3,82	2 1,73	3 0,73	1 0,29	0 0,16	143 133,36
3	89 83,76	3	11 7,88	19 20,00	16 22,88	16 19,41	17 13,55	7 8,21	1 4,46	0 1,62	1 0,44	1 0,31	0 0,31	89 102,38
4	70 81,03	4	6 4,56	6 10,71	15 13,99	13 13,38	13 10,43	9 7,00	8 4,18	5 2,27	1 1,14	2 0,53	1 0,40	79 68,59
5	63 71,78	5	3 1,90	5 5,09	3 7,50	9 8,01	9 6,91	8 5,10	0 3,33	3 1,97	2 1,07	1 0,54	0 0,45	43 41,87
6	60 59,70	6	1 0,73	1 2,21	4 3,63	4 4,29	3 4,07	1 3,29	3 2,34	1 1,50	1 0,88	1 0,48	1 0,45	21 23,87
7	47 47,33	7	0 0,26	0 0,89	2 1,62	2 2,11	2 2,19	3 1,92	0 1,48	1 1,02	0 0,64	1 0,38	0 0,40	11 12,91
8	35 36,14	8	0 0,09	0 0,34	0 0,67	1 0,96	1 1,08	0 1,03	0 0,85	1 0,64	0 0,43	0 0,27	0 0,32	3 6,68
9	28 26,78	9	0 0,03	1 0,12	2 0,26	0 0,41	1 0,50	3 0,51	0 0,46	0 0,37	0 0,26	0 0,18	0 0,24	7 3,34
≥ 10	70 62,11	≥ 10	0 0,02	0 0,07	1 0,15	0 0,26	0 0,35	1 0,41	0 0,43	1 0,57	0 0,32	0 0,24	0 0,41	3 3,03
		f(x ₁)	103 88,66	127 137,28	126 132,86	92 102,72	77 69,32	51 42,63	17 24,49	14 13,34	8 6,96	8 3,51	2 3,23	625 625,00

Table 9.7 The data set of Chen (1976), the NBD - GPHD model

with $\hat{\gamma} = 3,91$, $\hat{\theta} = 0,56$, $\hat{\alpha}_1 = -39,32$ and $\hat{\alpha}_2 = -38,867$.

s	f(s)	$x_2 \backslash x_1$	0	1	2	3	4	5	6	7	8	9	≥ 10	f(x_2)
0	23 25,22	0	23 25,22	28 27,77	17 18,96	5 10,24	6 4,79	1 2,02	1 0,79	0 0,29	0 0,10	1 0,03	0 0,01	86 90,22
1	64 55,22	1	36 27,45	56 39,45	38 32,01	11 20,49	10 11,11	7 5,35	2 2,35	0 0,96	0 0,37	0 0,13	0 0,08	140 138,75
2	76 75,93	2	23 18,52	31 31,63	28 31,19	27 23,16	15 14,34	11 7,79	2 3,82	2 1,73	3 0,73	1 0,29	0 0,16	143 133,36
3	89 83,76	3	11 9,88	19 20,00	16 22,88	16 19,41	17 13,55	7 8,21	1 4,46	0 2,22	1 1,02	1 0,44	0 0,31	89 102,38
4	70 81,03	4	6 4,66	6 10,71	15 13,99	13 13,38	13 10,43	9 7,00	8 4,18	5 2,22	1 1,14	2 0,53	1 0,40	79 68,59
5	63 71,78	5	3 1,90	5 5,09	3 7,50	9 8,01	9 6,91	8 5,10	0 3,33	3 1,97	2 1,07	1 0,54	0 0,45	43 41,87
6	60 59,70	6	1 0,73	1 2,21	4 3,63	4 4,29	3 4,07	1 3,29	3 2,34	1 1,50	1 0,88	1 0,48	1 0,45	21 23,87
7	47 47,33	7	0 0,26	0 0,89	2 1,62	2 2,11	2 2,19	3 1,92	0 1,48	1 1,92	0 0,64	1 0,38	0 0,40	11 12,91
8	35 36,14	8	0 0,09	0 0,34	0 0,67	1 0,96	1 1,08	0 1,03	0 0,85	1 0,64	0 0,43	0 0,27	0 0,32	3 6,68
9	28 26,78	9	0 0,03	1 0,12	2 0,26	0 0,41	1 0,50	3 0,51	0 0,46	0 0,37	0 0,22	0 0,18	0 0,24	7 3,34
≥ 10	75 62,11	≥ 10	0 0,02	0 0,07	1 0,15	0 0,26	0 0,35	1 0,41	0 0,43	1 0,37	0 0,32	0 0,24	0 0,41	3 3,03
		f(x_1)	103 88,66	127 137,28	126 132,86	92 102,72	77 69,32	51 42,63	17 24,49	14 13,34	8 6,96	8 3,51	2 3,23	625 625,00

9.3.1 The NBD - GPHD model

The parameters for the NBD-GPHD model were estimated using the method of moments. The theoretical distribution was fitted with the parameters

$$\hat{\gamma} = 3,91, \quad \hat{\theta} = 0,56, \quad \hat{\alpha}_1 = -39,32 \quad \text{and} \quad \hat{\alpha}_2 = -38,867.$$

In Table 9.7 the empirical and the theoretical frequencies obtained by fitting the model are presented.

From the χ^2 test performed for the data we obtained the results summarized in Table 9.8.

Table 9.8 The χ^2 results obtained from Table 9.7

	χ^2	d.f.	$P(\chi^2)$
Bivariate table	35,56	41	0,711
x_1 marginal	11,10	5	0,049
x_2 marginal	8,99	5	0,110
Third marginal	9,12	13	0,764

The graphical displays of the x_1 marginal, x_2 marginal and the third marginal are presented in Figures 9.11, 9.12 and 9.13 respectively.

Figure 9.11 The x_1 marginal, data of Chen (1976),
the NBD - GPHD model

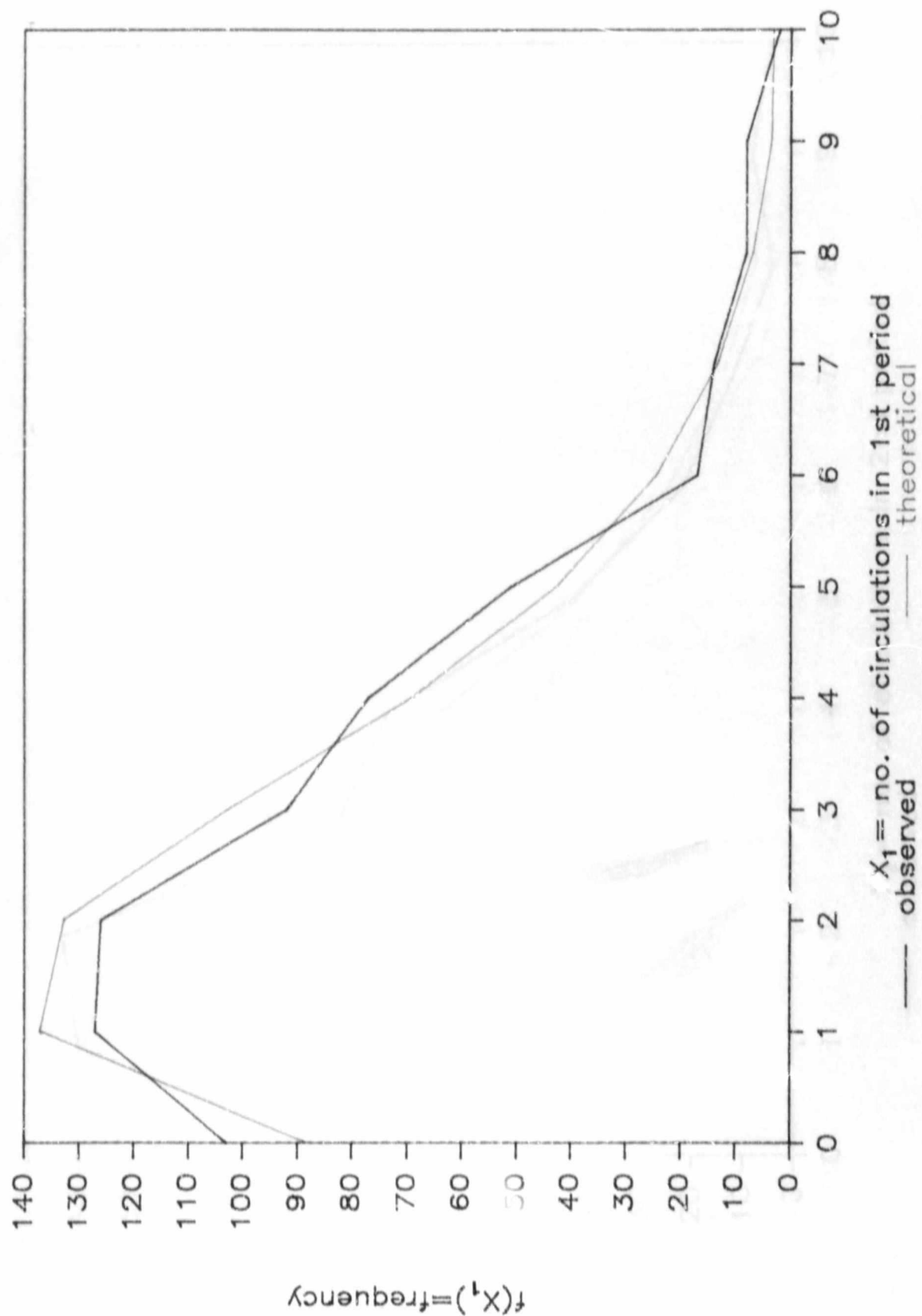


Figure 9.12 The x_2 marginal, data of Chen (1976),
the NBD - GPHD model

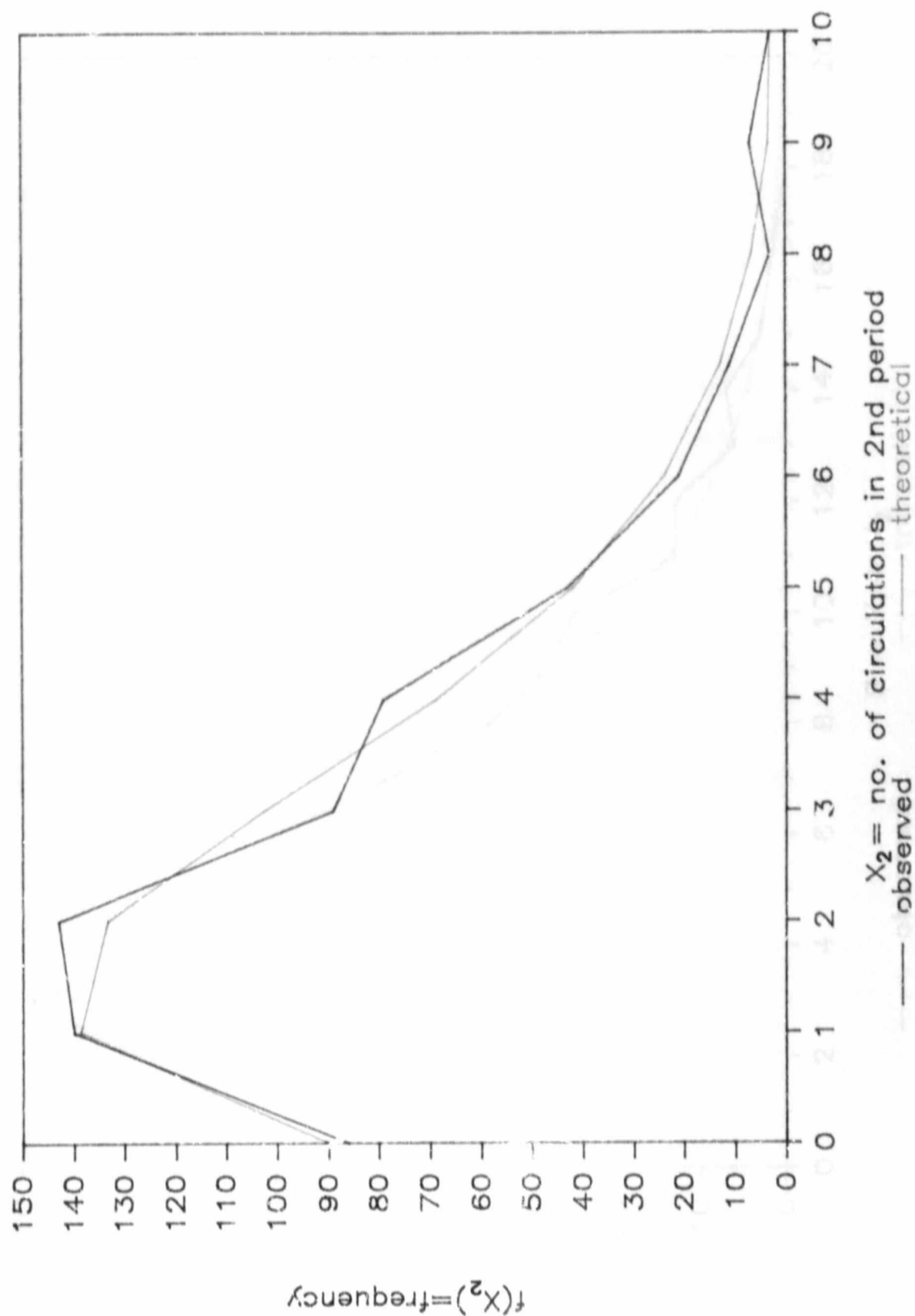


Figure 9.13 The third marginal, data of Chen (1976),
the NBD - GPHD model

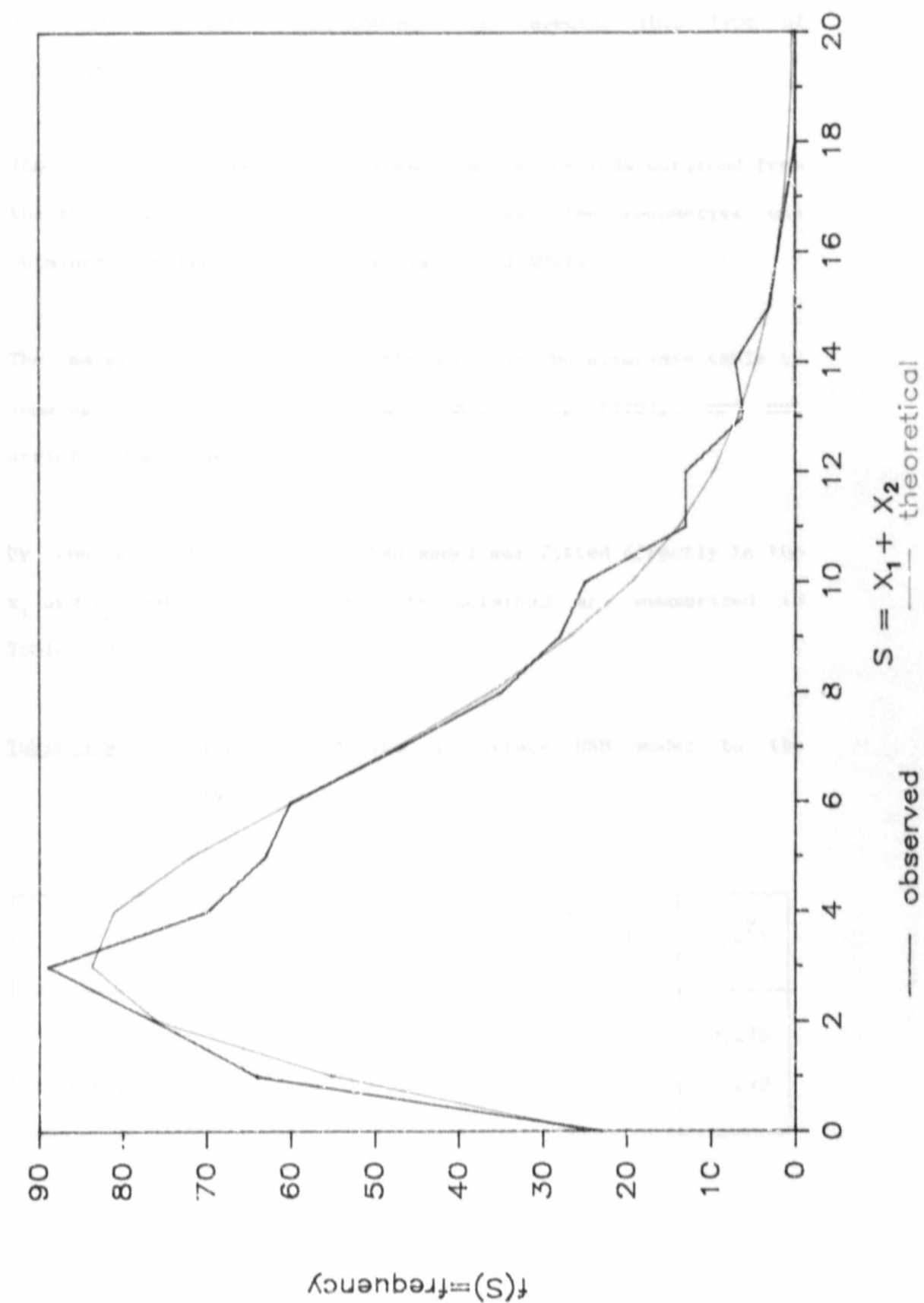


Figure 9.14 shows the graphical display of the empirical and the theoretical regression curve $E(x_2|x_1)$. The regression curve is of a concave shape with a strong tendency towards a straight line. For negative values of α_1 and α_2 we expected this type of curvilinear regression.

The correlation coefficient between the two periods obtained from the theoretical fit is $\rho = 0,42411$ and the respective one obtained from the empirical data is $\rho = 0,40279$.

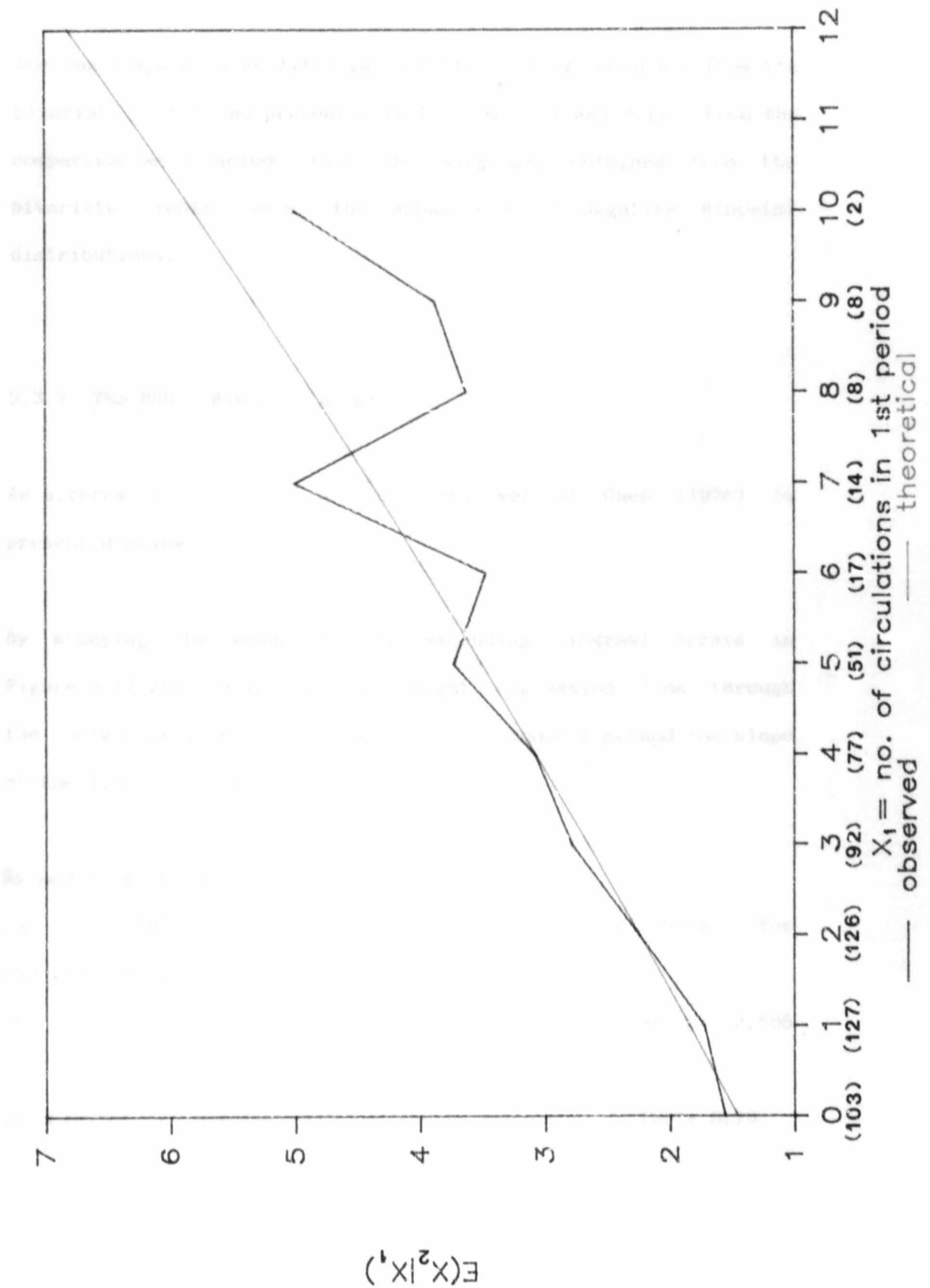
The marginals x_1 and x_2 , obtained from the bivariate table by summing the vertical and the horizontal respectively, are not strictly distributed NBD.

By contrast, the univariate NBD model was fitted directly to the x_1 and x_2 marginals. The results obtained are summarized in Table 9.9.

Table 9.9 The fit of the univariate NBD model to the marginals

	$\hat{\gamma}$	$\hat{\theta}$	χ^2	d.f.	$P(\chi^2)$
x_1 marginal	3,64	0,409	8,70	7	0,275
x_2 marginal	4,48	0,355	9,14	7	0,243

Figure 9.14 The regression curve $E(x_2|x_1)$, Chen (1976),
the NBD - GPHD model



The observed and the theoretical fittings of the univariate NBD model are presented for the x_1 and the x_2 marginals in Figures 9.15 and 9.16 respectively.

One can compare these fittings with the fittings obtained from the bivariate table and presented in Figures 9.11 and 9.12. From the comparison we conclude that the marginals obtained from the bivariate table have the appearance of Negative Binomial distributions.

9.3.2 The NBD - Binomial model

An alternative analysis of the data set of Chen (1976) is presented below.

By studying the means of the ascending diagonal arrays in Figure 9.17 one can see that a straight regression line through the origin is suitable. Using the least squares method the slope of the line was estimated as $\hat{p} = 0,505$.

By applying the bivariate model with:

(a) the NBD with the parameters $\hat{\gamma} = 3,91$ and $\hat{\theta} = 0,56$ for the third marginal and

(b) the Binomial distribution with the parameter $\hat{p} = 0,505$ for the ascending diagonal arrays,

we obtained the theoretical frequencies presented in Table 9.10.

Figure 9.15 The x_1 marginal, data of Chen (1976),
the univariate NBD model

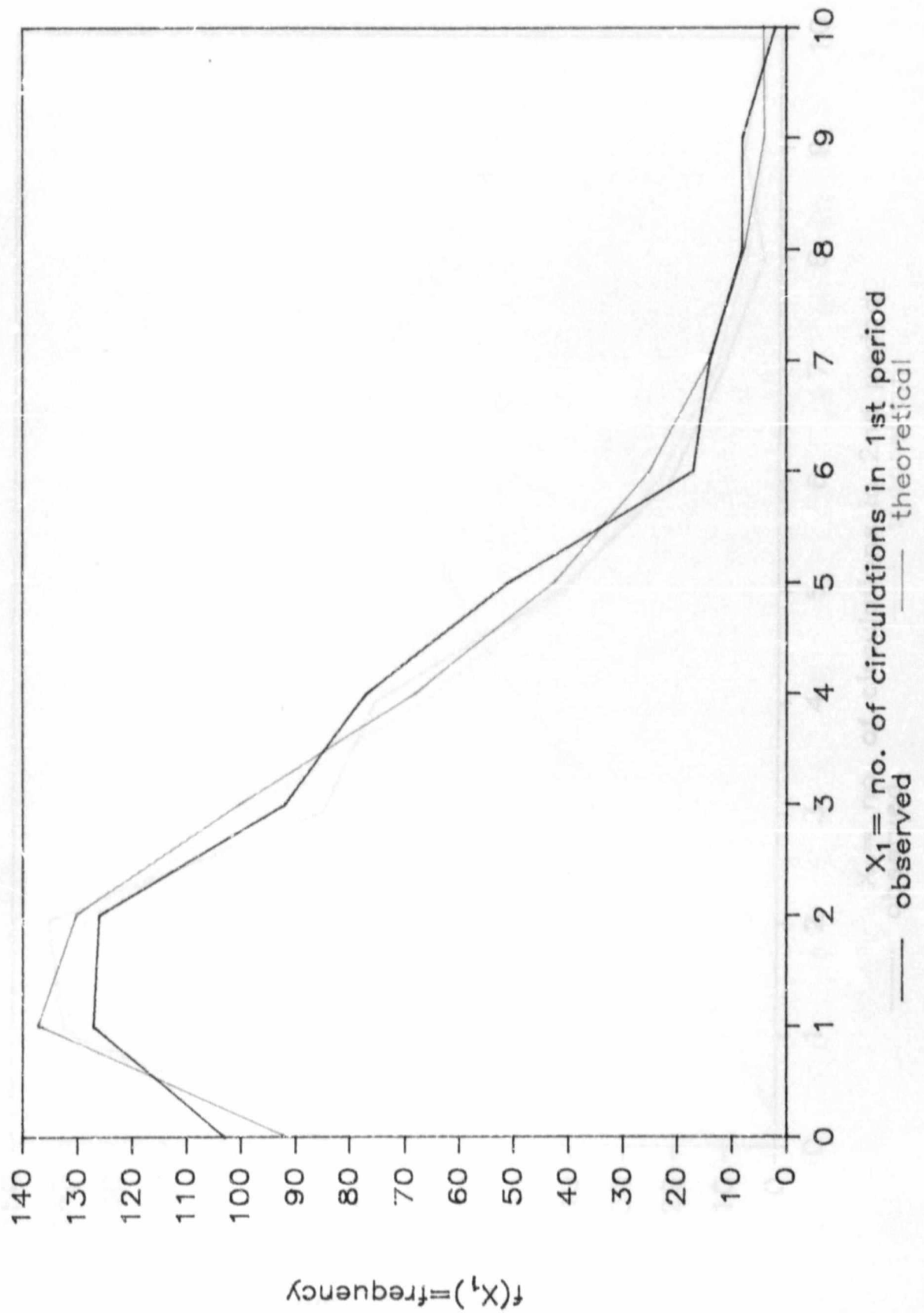


Figure 3.16 The x_2 marginal, data of Chen (1976),
the univariate NBD model

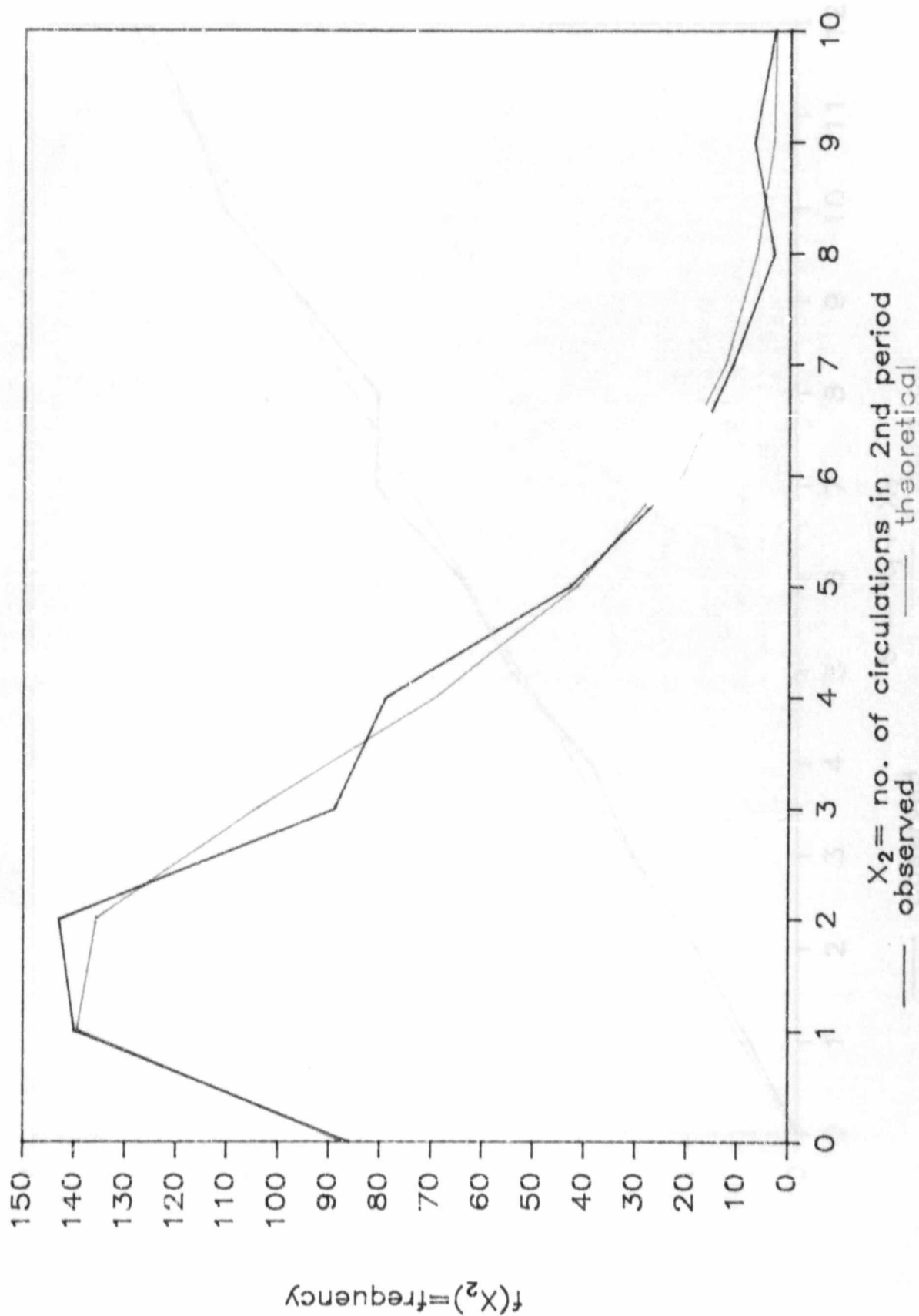


Figure 9.17 The means of the ascending diagonal arrays
 ($\hat{p} = 0,505$), Chen (1976)

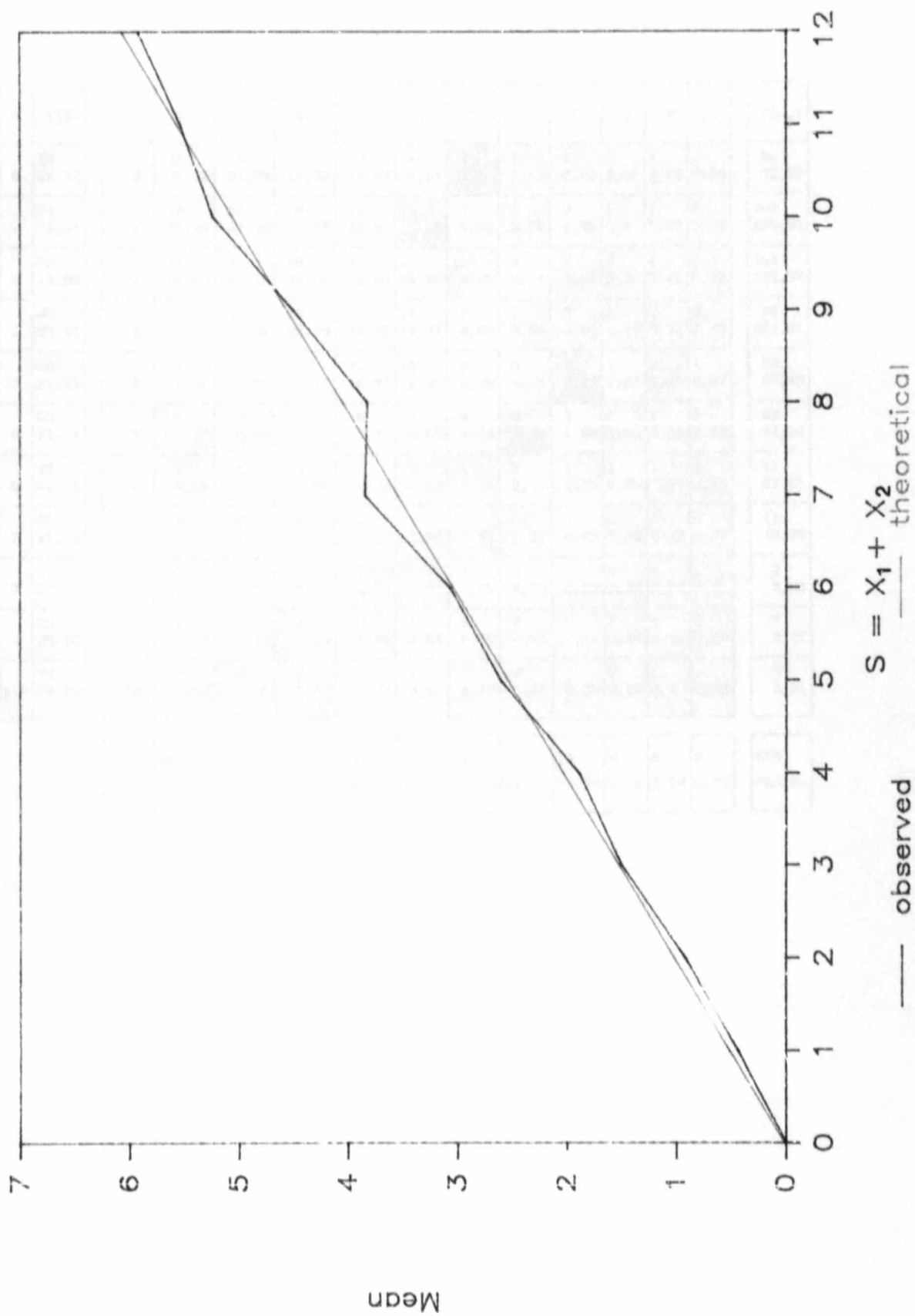


Table 9.10 The data set of Chen (1976), the NBD - Binomial model

with $\hat{\gamma} = 3,91$, $\hat{\theta} = 0,56$ and $\hat{p} = 0,505$

s	f(s)	$x_2 \backslash x_1$	0	1	2	3	4	5	6	7	8	9	≥ 10	f(x ₂)
0	23 25,22	0	23 25,22	28 27,89	17 19,36	9 10,79	6 5,27	1 2,36	1 0,99	0 0,40	0 0,15	1 0,06	0 0,03	86 92,52
1	64 55,23	1	36 27,34	36 37,96	38 31,72	11 20,6	10 11,66	7 5,82	2 2,72	0 1,20	0 0,51	0 0,21	0 0,12	140 139,82
2	76 75,92	2	23 18,60	31 31,10	28 30,38	27 22,65	15 14,27	11 8,00	2 4,11	2 1,98	3 0,90	1 0,40	0 0,28	143 132,67
3	89 83,77	3	11 10,16	19 19,85	16 22,20	16 18,65	17 13,07	7 8,06	1 4,53	0 2,36	1 1,16	1 0,54	0 0,43	89 101,01
4	70 81,02	4	6 4,86	6 10,88	15 13,71	13 12,81	13 9,88	9 6,66	8 4,05	5 2,28	1 1,20	2 0,60	1 0,52	79 67,45
5	63 71,78	5	3 2,13	5 5,38	3 7,53	9 7,75	9 6,52	8 4,76	0 3,12	3 1,88	2 1,06	1 0,56	0 0,55	43 41,24
6	60 59,70	6	1 0,88	1 2,46	4 3,80	4 4,26	3 3,89	1 3,06	3 2,15	1 1,38	1 0,83	1 0,47	1 0,49	21 23,67
7	47 47,33	7	0 0,34	0 1,06	2 1,79	2 2,18	2 2,14	3 1,81	0 1,36	1 0,93	0 0,53	1 0,15	0 0,40	11 12,95
8	35 36,14	8	0 0,13	0 0,44	0 0,80	1 1,05	1 1,11	0 1,00	0 0,79	1 0,58	0 0,38	0 0,24	0 0,31	83
9	28 26,80	9	0 0,05	1 0,18	2 0,34	0 0,48	1 0,54	3 0,52	0 0,44	0 0,34	0 0,24	0 0,16	0 0,20	7 3,19
≥ 10	70 62,09	≥ 10	0 0,04	0 0,10	1 0,25	0 0,37	0 0,45	1 0,48	0 0,45	1 0,35	0 0,28	0 0,19	0 0,35	3 3,35

f(x ₁)	103	127	126	92	77	51	17	14	8	8	2	625
	89,75	137,30	131,88	101,65	68,70	42,53	24,71	13,69	7,30	3,76	3,71	625,00

From the χ^2 test performed on the data we obtained the results summarized in Table 9.11.

Table 9.11 The χ^2 results obtained from Table 9.10

	χ^2	d.f.	$P(\chi^2)$
Bivariate table	38,62	43	0,662
x_1 marginal	9,92	6	0,128
x_2 marginal	8,95	6	0,176
Third marginal	8,78	12	0,722

The graphical displays of the x_1 marginal, the x_2 marginal and the third marginal for the NBD - Binomial model are presented in Figures 9.18, 9.19 and 9.20 respectively.

Figure 9.21 shows the graphical display of the empirical and the theoretical regression curve $E(x_2|x_1)$.

The correlation coefficient between the variables x_1 and x_2 obtained from the theoretical fit is $\rho = 0,3888$.

A comparison between the NBD - GPHD model and the NBD - Binomial model leads to the conclusion that the latter is more suitable for the data set of Chen (1976).

Figure 9.18 The x_1 marginal, data of Chen (1976),
the NBD - Binomial model

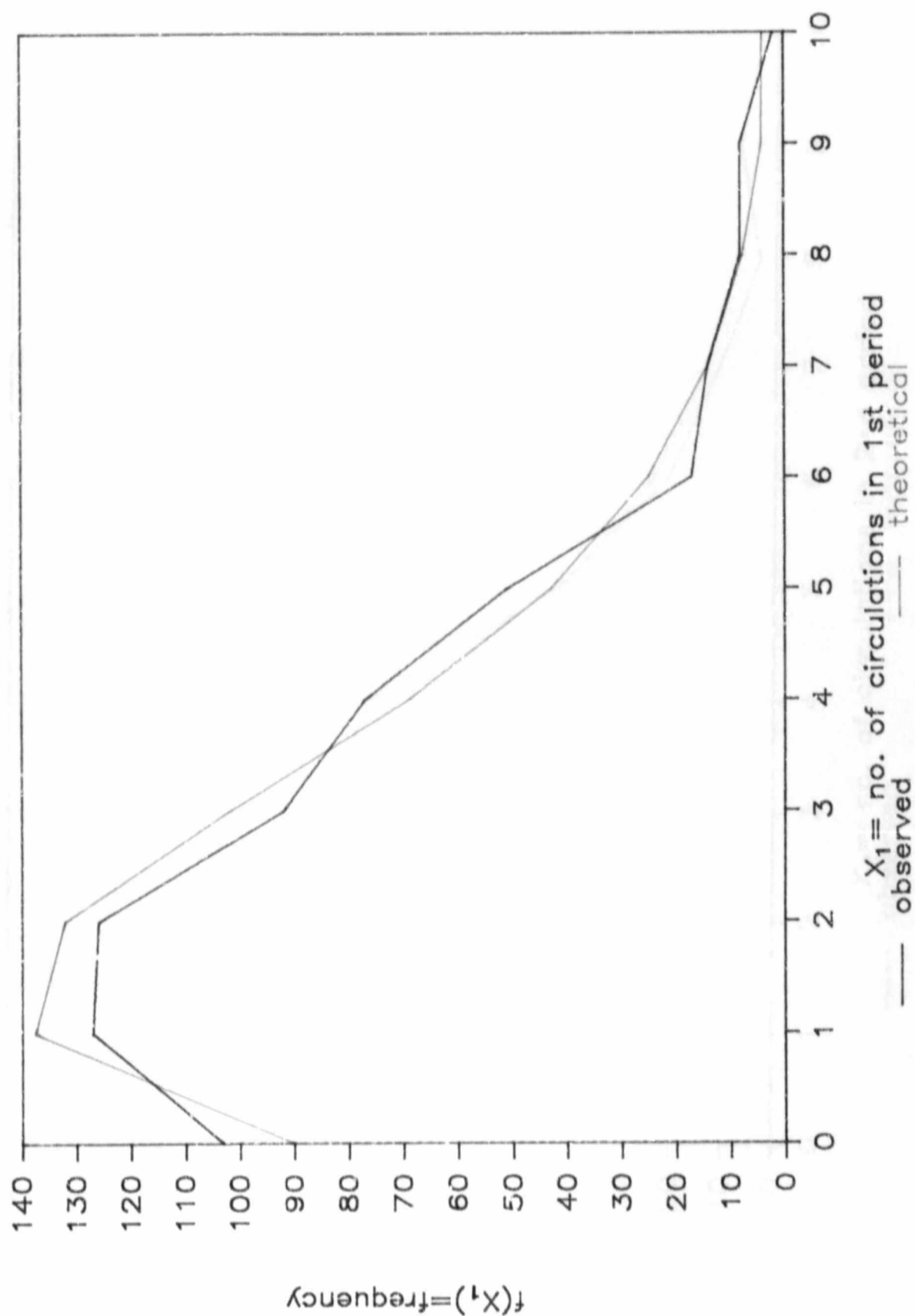


Figure 9.19 The X_2 marginal, data of Chen (1976),
the NB - Binomial model

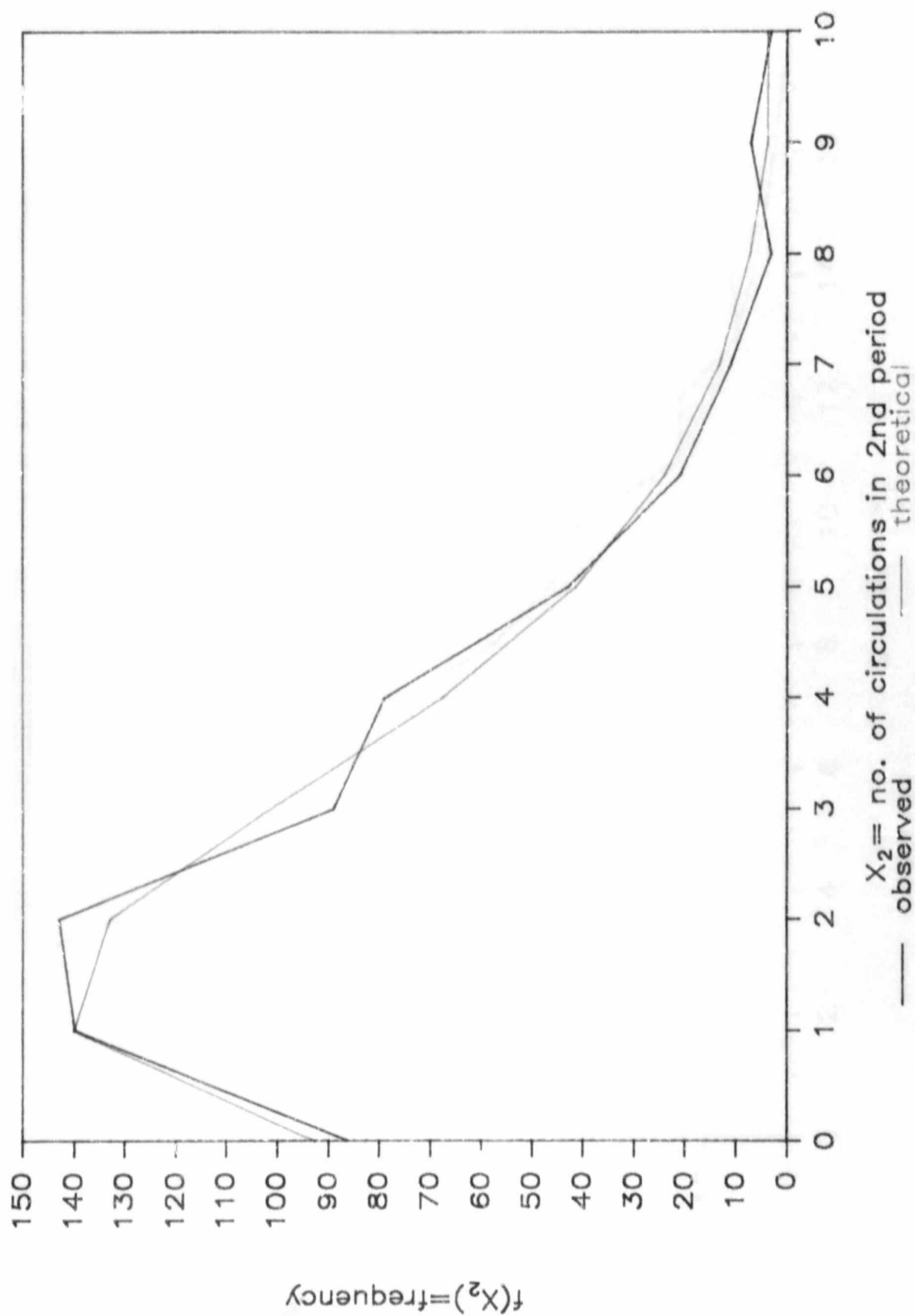


Figure 9.20 The third marginal, data of Chen (1976),
the NBD - Binomial model

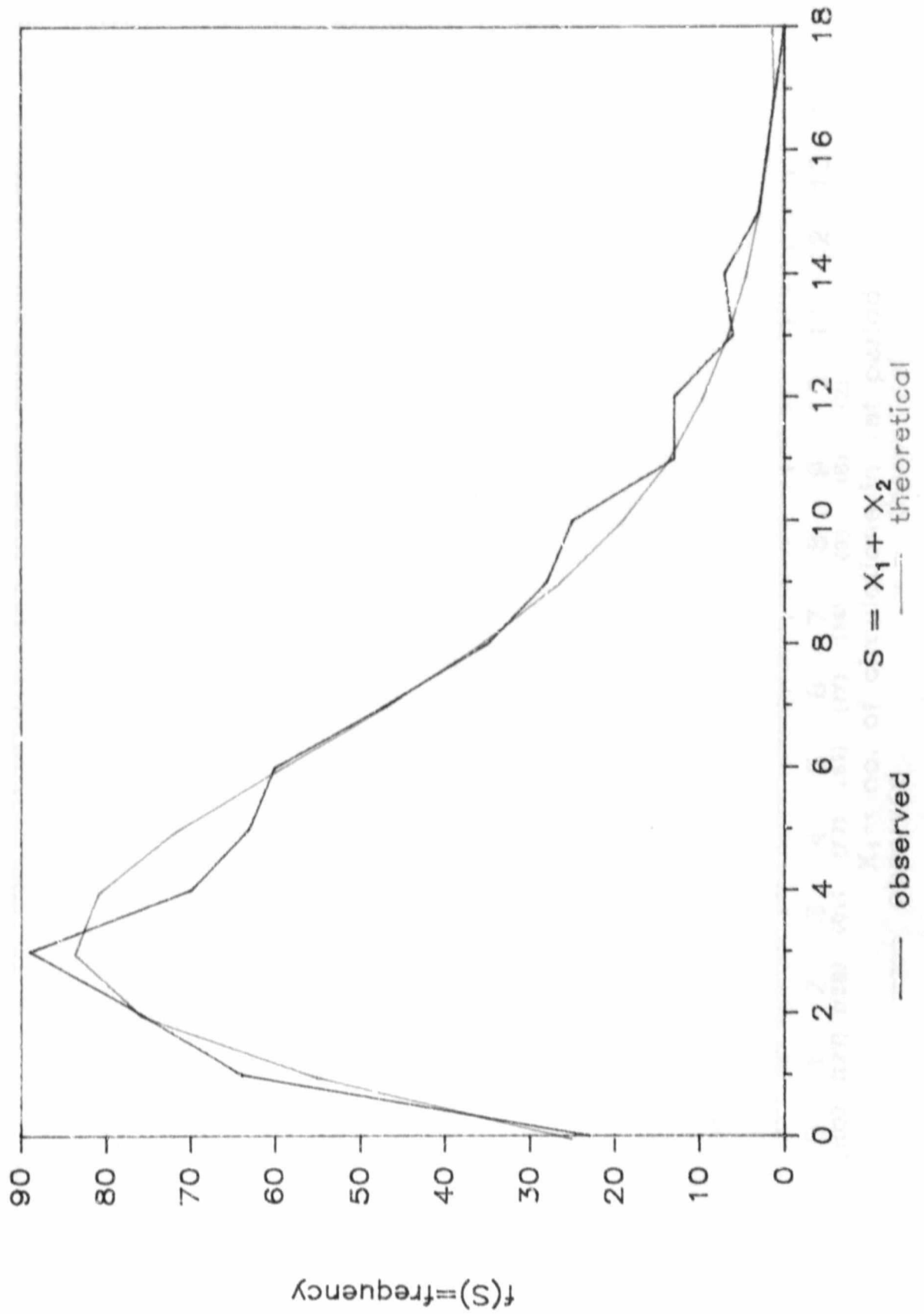
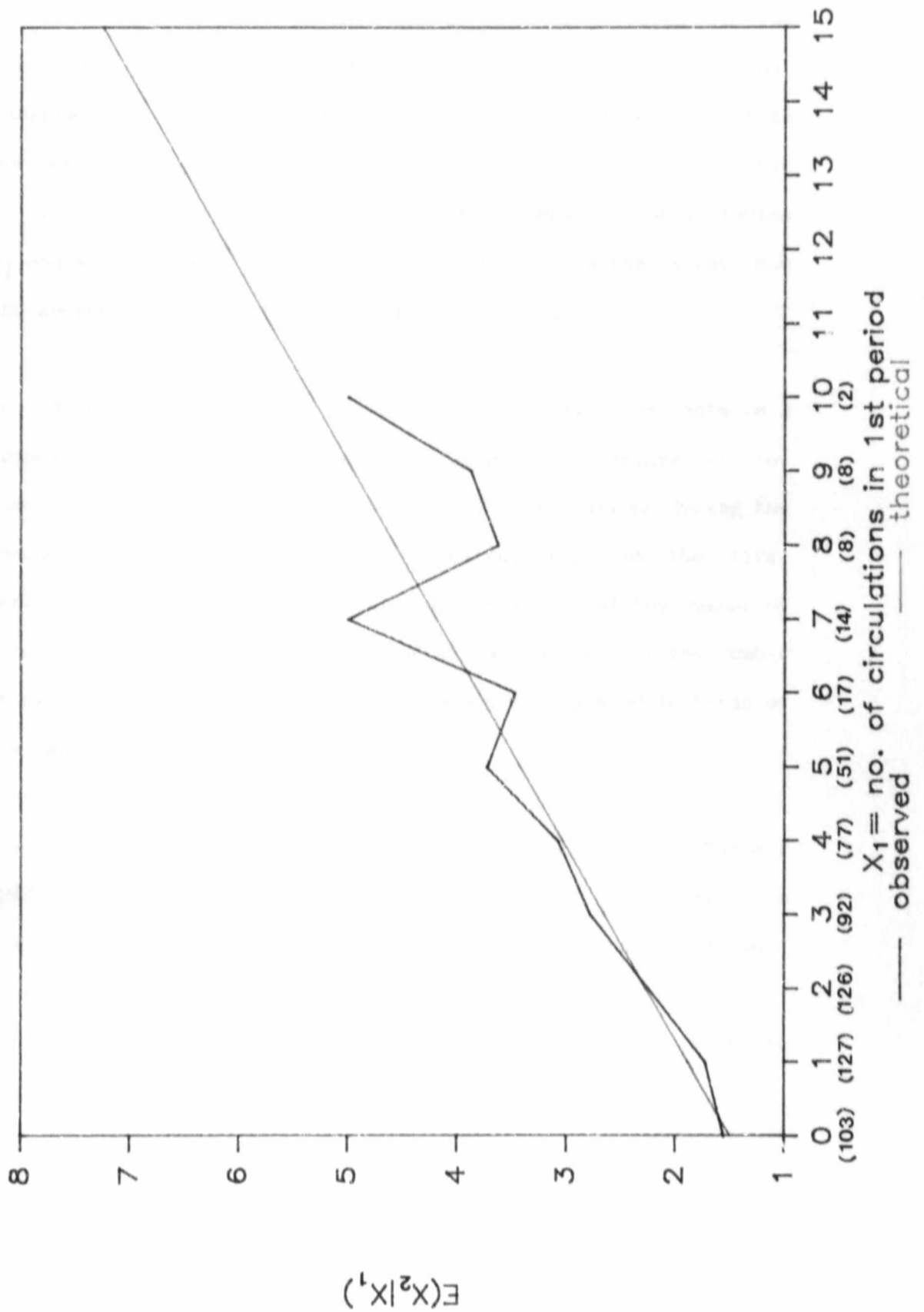


Figure 9.21 The regression curve $E(x_2|x_1)$, Chen (1976),
the NBD - Binomial model



9.4 The Data Set from the Library Circulation Field,
used by Cane and Burrell (1982)

The data in Table 9.12 represent the circulation of items in two successive time periods in the University of Sussex "short loan" library. All the copies which were available throughout both periods of 1975-76 and the entire academic year 1976-77 were taken in consideration, a total of 28440 volumes. The variables x_1 and x_2 represent the number of circulations in the first and the second periods respectively and $S = x_1 + x_2$.

In their paper Cane and Burrell (1982) gave the data in a truncated form. They presented some graphical displays of the conditional frequency distributions of circulations during the second period, given the number of circulations in the first period. The authors exhibit the conditional mean of the number of circulations $E(x_2|x_1)$ and the conditional variance of the number of circulations $V(x_2|x_1)$. Their graphs are displayed in terms of the empirical data.

Since the number of observations presented by Cane and Burrell (1982) is unusually large, we have an excellent opportunity to demonstrate the fit of the model developed in this investigation.

Burrell (1986) was kind enough to supply this researcher with the untruncated set of observations.

Author Barr Aiala

Name of thesis New Statistical Models For Discrete Uni- And Multivariate Data Sets With Special Reference To The Dirichlet Multinomial Distribution. 1987

PUBLISHER:

University of the Witwatersrand, Johannesburg

©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.