*Dairying*

UNIVERSITY OF MANCHESTER

FACULTY OF ECONOMIC AND SOCIAL STUDIES

DEPARTMENT OF AGRICULTURAL ECONOMICS

# STRUCTURAL CHANGE IN THE DAIRY INDUSTRY IN ENGLAND AND WALES

## AN APPLICATION OF MARKOV CHAIN ANALYSIS

DAVID COLMAN
AND
DENNIS LEECH

BULLETIN 125/M24

*Price:* FIVE SHILLINGS

UNIVERSITY OF MANCHESTER

FACULTY OF ECONOMIC AND SOCIAL STUDIES

*With the Compliments*

*of the*

AGRICULTURAL ECONOMICS DEPARTMENT
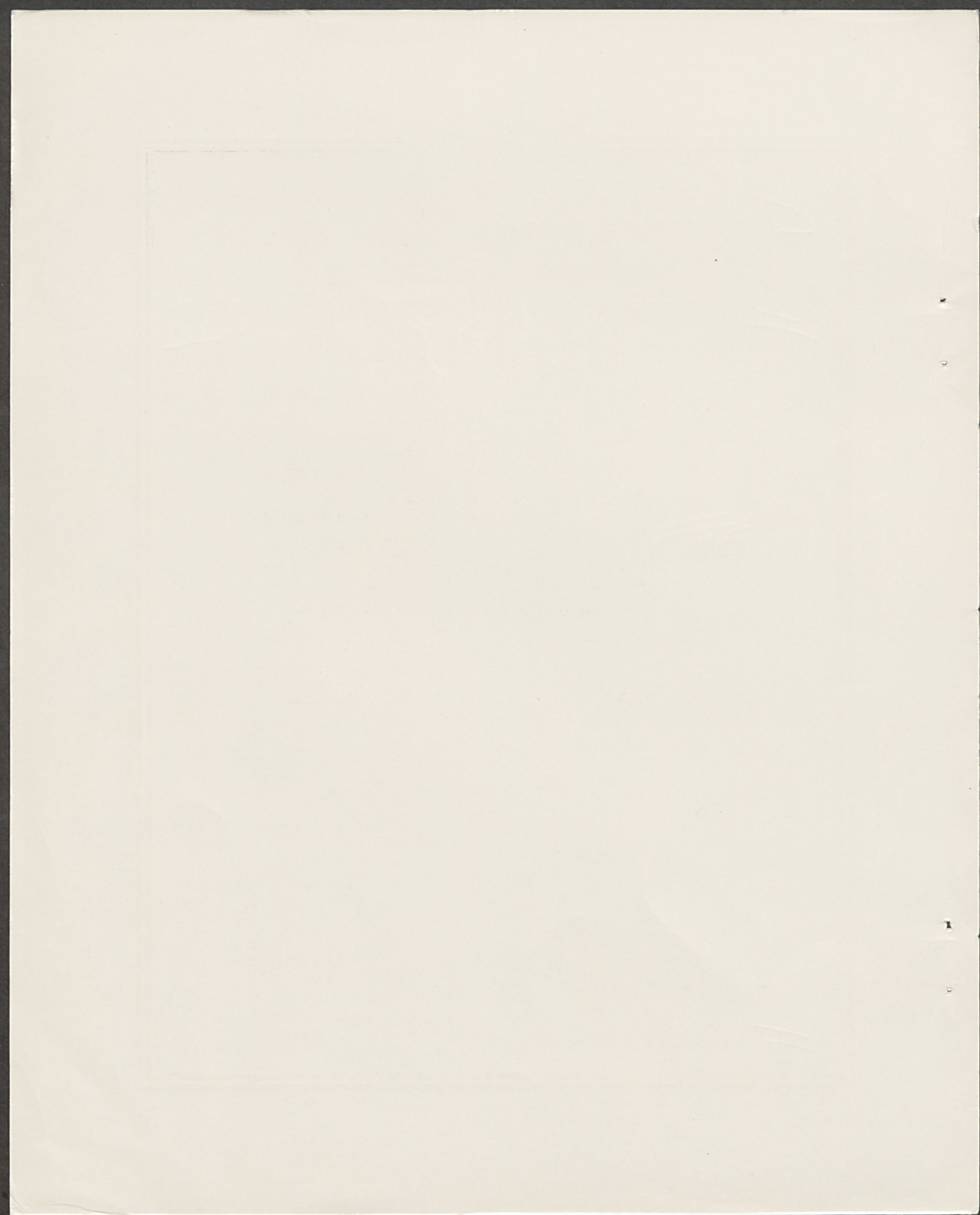
STRUCTURAL CHANGE

IN THE DAIRY INDUSTRY

IN ENGLAND AND WALES

An Application of Markov Chain Analysis


David Colman
and
Dennis Leech


January 1969

# CONTENTS

## Introduction

In recent years the dairy industry has undergone considerable change in its structure as regards the number and size of herds and their geographical location[1]. While the number of milk producers in England and Wales has fallen sharply (by 9.3% between March 1963 and March 1965) annual milk production has remained fairly constant, and has even risen by a small amount. This has been due to the increased average output of those producers who have remained in production. The result has been that the industry has become more concentrated. Milk production has increasingly become the concern of a smaller number of larger enterprises.

The regional structure of dairy farming has likewise shown noticeable changes. Production has increasingly been carried out in the MMB regions bordering on the Irish Sea (i.e. the Northern, Northwestern, North Wales and South Wales Regions). These four regions were the only ones to increase production from 1962 to 1965 despite the considerable fall in producer numbers which occurred in these, as well as all other regions.

The overall pattern, therefore, has been a shifting distribution of milk production towards larger farms and towards the west coastal regions.

The present study is an attempt to apply the technique of Markov chains to the problem of predicting the future structure and output of

---

1. The following description is drawn from "Changes in Milk Output 1963 to 1965" published by the MMB in 1966.

the dairy industry, at both the national and regional levels. The regional analysis is of interest for comparative purposes in that it enables examination of regional differences in the pattern of structural change both during the sample period and into the predicted future. It is hoped that prediction using regional Markov chains will lead to accurate and detailed results nationally by taking into account small component effects which would be overlooked by more conventional, aggregative methods.

The data are taken from the MMB's Permanent Producer Sample for the years 1963 to 1967. This is a random sample of more than 5,000 producers who were asked to keep detailed records of their milk output for every month since April 1963. The sample is stratified by regions and by output of milk, which means that all regions and all sizes of herds are represented. The quality of this sample enables inferences about the behaviour of the population of dairy farmers to be made with a fair degree of confidence on the basis of sample observations.

A further objective of the study is an assessment of the validity of Markov chain analysis in the light of these very good data. In particular it is of interest to know whether the dairy industry's pattern of change, as indicated by the sample, is substantially the same from year to year. The answer to this question is fundamental in determining the suitability of Markov chain analysis as a predictive tool. If the year to year variations in this pattern are large it is obviously not possible to make predictions with confidence. In this methodoligical respect the current study follows

up an earlier one[1] which used data from a relatively poorly designed
sample of dairy farms in the northwestern region. This sample, based on
the Farm Survey, was not randomised and was not fully representative
of the industry in the region, two faults which should have been overcome
by the Permanent Producer Survey. The lack of success of the earlier
study in making predictions could have been due either to the
inapplicability of the Markov chains technique to the dairy industry, or
to errors in the sample data. The latter problem is presumed to be minimsed
in the data used in the current study and for this reason it is hoped
that it will be possible to evaluate the usefulness of the Markov chains
technique when applied to the dairy industry.

1. D. R. Colman "Application of Markov chain analysis to structural change
in the North West dairy industry". Journal of Agricultural Economics
Vol. XVIII No. 3 1967

## Outline of the Study

The line of approach followed by the study is first of all to derive values of the parameters upon which prediction by Markov chains depends, i.e. the transition probabilities matrix for each region. A transition probabilities matrix, explained in more detail below, characterises the pattern of change in the size distribution[1] of milk producers which took place during the period embraced by the sample. Comparison of the transition probabilities matrices for different regions reveals regional peculiarities in the pattern of change.

The next step is to obtain a prediction of the future size distribution for each region. This consists of the predicted numbers of producers in each of the six classes. From these regional figures the future national size distribution is arrived at by aggregating over all eleven regions.

---

1. Producers are here classified by output according to the six size classes used by the MMB in "The Structure of Dairy Farming" and "Changes in Milk Output". The classes are defined below with respect to annual gallonage:

| | |
|---|---|
| "Very Small" | under 7,500 |
| "Small" | 7,500 - 14,999 |
| "Average" | 15,000 - 24,999 |
| "Substantial" | 25,000 - 49,999 |
| "Large" | 50,000 - 99,999 |
| "Very Large" | 100,000 - and over |

There is also a further class, denoted by 0, to represent the case of no milk production in a particular year. This is necessary because some producers are observed to enter the industry and some existing producers are observed to leave

From the sample data average output per producer is found for each of the six size classes in each region. From these, together with the predictions of the number of producers in each class, predicted output figures are arrived at for each size class and each region. Regional output is then found by aggregating over classes and national output found by aggregating over the regions.

As regards the methodology of Markov chain analysis, it is of importance to determine if the components of each transition probabilities maxtrix remained fairly constant throughout the period covered by the sample. To answer this question a statistical significance test (the chi-square test) is carried out.

## The Pattern of Structural Change

## The Transition Probabilities Matrix

It was stated in the previous section that a regional transition probabilities matrix characteristizes the pattern of structural change in a region. As an illustration and by way of explanation, the transition probabilities matrix for the South Wales region is presented below

### TABLE I  TRANSITION PROBABILITIES MATRIX FOR SOUTH WALES

| Class in First Year | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9975 | .0018 | .0003 | .0002 | .0002 | 0 | 0 |
| Very Small | .1250 | .7780 | .0896 | .0075 | 0 | 0 | 0 |
| Small | .0217 | .0924 | .8007 | .0815 | .0036 | 0 | 0 |
| Average | .0153 | .0122 | .0948 | .7645 | .1131 | 0 | 0 |
| Substantial | .0217 | 0 | .0072 | .0797 | .8116 | .0797 | 0 |
| Large | .0114 | 0 | 0 | .0277 | .0568 | .8636 | .0455 |
| Very Large | 0 | 0 | 0 | 0 | 0 | 0 | 1.0 |

Each of the entries in this matrix is the probability associated with a milk producer taking a particular course of action between one year and the next. Possible courses of action are, for example, increasing or decreasing the size of output, maintaining the same output, or ceasing production entirely. In the matrix the classes referring to the initial year are represented by the rows , and those for the following year by the columns. There are seven rows and seven columns, there being six size classes of producers, and one class to represent those not currently in production but who may either be so at some future date or have produced in the past.

As an illustration of the meaning of the transition probabilities matrix consider the row labelled "Average". It gives the probability distribution of outcomes for one year, given that a producer was in the Average class at the beginning of the year. Thus reading along the row gives the probabilities of all the respective outcomes conditional upon starting in the Average class. The probability that a producer leaves the dairy industry (column 0) is 0.0153. Similarly the probabilities of reducing output to the level of the Very Small or Small classes are, respectively, 0.0122 and 0.0948. The combined probability of a decrease in the size of output, conditional upon beginning in the Average class, is then the sum of these three values:

$$0.0153 + 0.0122 + 0.0948 = 0.1223$$

The entry in the diagonal, 0.7645, is the probability that the producer remains in the same class. The probability of the producer increasing his output from the "Average" level is 0.1131, that is, the probability of entering

the "Substantial" class. The probabilities of moving into the "Large" or "Very Large" classes are both zero, indicating that the possibility of a move of this type is excluded from the model.

It will be seen from this that a producer in the "Average" class is slightly more likely to decrease the level of his output than to raise it (a probability of 0.1223 as against 0.1131). However, the greatest likelihood is that he will approximately maintain his level of output and remain in the "Average" class (a probability of 0.7645).

The elements in a single row of the matrix must have the property of summing to unity. This is so because each row is a probability distribution and describes all the possible outcomes for a given starting class. The columns of the matrix do not have this property.

As a purely hypothetical illustration of the meaning of a transition probabilities matrix, suppose there were 100 milk producers in the Average class in South Wales in 1966-7. By the following year 1967-8, they would be expected to be distributed according to the transition probabilities in row 3 of the matrix. Thus, (to the nearest whole numbers) the distribution would be:

> 2 producers ceased production altogether
>
> 1 producer reduced output to Very Small class
>
> 9 producers reduced output to Small class
>
> 76 producers remaining in Average class
>
> 11 producers increased output to Substantial class
>
> No producers moved to Large or Very Large classes.

A similar interpretation to this one attaches to each row of the matrix

Thus it will be seen that the probability of a producer in the Very Large class remaining there is unity. This means that no producer in South Wales who is included in the Very Large class is ever expected to leave it.[1]

The row of the matrix denoted by O is the probability distribution of potential entrants to dairying. The probability that a potential dairy farmer will <u>not</u> actually take up production during the year is 0.9975[2]. The combined probability of his entering the industry is 0.0025. Hence it is highly unlikely that a potential entrant will, in fact, begin production of milk.

Derivation of the Transition Probabilities Matrices

The Permanent Producer Survey, which supplied the data for the present study, comprised a large sample of milk producers in each MMB region. These producers were requested to keep records of their gallonage for the four years from 1963-4 to 1966-7, and from the information thus supplied the transition probabilities MATRICES ~~notices~~ were derived.

---

1. This restrictive condition derives from the fact that in South Wales in 1963-4 there were 5 producers in the Very Large class all of whom were included in the sample and none of whom left the class during the sample period. The derivation of the transition probabilities matrices is treated in the next section and Appendix I.

2. Details of how this result was obtained are presented in Appendix I section 2

The first step was to allocate each producer to one of the six classs[1] according to size, on the basis of recorded output in 1963-4. The allocation was repeated using the records of the following year's output. Hence the career of each producer in the sample was known for this two year period.

For example, a producer may have been in the Average class in 1963-4 and the following year may have ceased production. Alternatively, he may have remained in the Average class, and so on.

From this information a table was drawn up each of whose entries is the total number of sample producers following a particular course of action. The table (or matrix) for the South Wales region is presented below

TABLE 2

| | 1964-5 Class | | | | | | |
|---|---|---|---|---|---|---|---|
| 1963-4 Class | 0 | V. Small | Small | Average | Substantial | Large | V. Large |
| 0 | 2010 | 3 | 1 | 0 | 0 | 0 | 0 |
| V. Small | 25 | 157 | 14 | 3 | 0 | 0 | 0 |
| Small | 3 | 16 | 158 | 15 | 0 | 0 | 0 |
| Average | 2 | 3 | 10 | 33 | 16 | 0 | 0 |
| Substantial | 2 | 0 | 0 | 3 | 28 | 6 | 0 |
| Large | 1 | 0 | 0 | 0 | 4 | 23 | 2 |
| V. Large | 0 | 0 | 0 | 0 | 0 | 0 | 5 |

1. See p. 3 for a definition of the classes

The rows of this matrix represent the classes to which producers belonged in 1963-4 and the columns those to which they were allocated a year later in 1964-5. For example, the entry in the row labelled "Average" and the column labelled "Substantial" contains the information that, of those producers assigned to the "Average" Class in 1963-4, 16 had by the following year increased their output and been included in the "Substantial" class. Similarly, 83 of them maintained the same output and remained in the "Average" class.

The Survey period was 4 years and therefore, for each region, there are 3 such matrices as the one above. The transition probabilities matrix was calculated from the sum of these. This "summation matrix" is presented for the South Wales region, below (Table 3)

## TABLE 3

|  | 0 | V. Small | Small | Average | Substantial | Large | V. Large | Row Total |
|---|---|---|---|---|---|---|---|---|
| 0 | 6095 | 11 | 2 | 1 | 1 | 0 | 0 | 6110 |
| V. Small | 67 | 417 | 48 | 4 | 0 | 0 | 0 | 536 |
| Small | 12 | 51 | 442 | 45 | 2 | 0 | 0 | 552 |
| Average | 5 | 4 | 31 | 250 | 37 | 0 | 0 | 327 |
| Substantial | 3 | 0 | 1 | 11 | 112 | 11 | 0 | 138 |
| Large | 1 | 0 | 0 | 2 | 5 | 76 | 4 | 88 |
| V. Large | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 19 |

Each element in the transition probabilities matrix for South Wales was then calculated as the ratio of each entry to its corresponding row total. For example, the elements in the Average row are presented:

| | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
|---|---|---|---|---|---|---|---|
| Average | 5/327 | 4/327 | 31/327 | 250/327 | 37/327 | 0 | 0 |
| Average | .0153 | .0122 | .0948 | .7645 | .1131 | 0 | 0 |

By the same procedure the elements of each row of the transition probabilities matrix were derived and the matrix built up row by row.

## 4 Regional Differences in Industrial Stability

### (1) Stability of regional dairy industries

The degree of stability of a regional dairy industry is indicated by the elements of its transition probabilities matrix in two ways.

The first of these is the size of the probabilities in the diagonal. Clearly, the closer each is to one the more stable is the industry. If all the diagonal elements were one, there would be perfect stability, with no changes in industrial structure, all producers continuing to produce roughly the same output each year.

The second indicator of stability is the extent to which the off-diagonal probabilities are dispersed. If there is little dispersion it is an indication of a degree of stability, the only movement between classes taking place between adjoining ones. If, however, there is a wide spread

of the off-diagonal probabilities, including the possibility of, say, a producer moving from the Large class to 0 (no production), then there are grounds for saying that the industry is unstable and subject to rapid structural changes.

TABLE 4

TRANSITION PROBABILITIES MATRIX FOR THE SOUTHEASTERN REGION

| Class in First Year | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | V. Small | Small | Average | Substantial | Large | V. Large |
| 0 | .9900 | .0030 | .0011 | .0017 | .0011 | .0011 | 0 |
| V. Small | .4086 | .5054 | .0430 | .0215 | .0108 | .0108 | 0 |
| Small | .0738 | .1393 | .6475 | .1230 | .0164 | 0 | 0 |
| Average | .0752 | .0451 | .1203 | .6541 | .1053 | 0 | 0 |
| Substantial | .0335 | .0239 | .0287 | .0670 | .7512 | .0957 | 0 |
| Large | .0079 | .0236 | .0079 | .0236 | .0945 | .7402 | .1024 |
| V. Large | 0 | 0 | 0 | 0 | .0045 | .0315 | .9640 |

By way of comparison of two transition probabilities matrices, consider those for the South Wales (Table 1) and South Eastern regions (Table 4). (These regions were found to exhibit the greatest stability and instability respectively of all the 11 regions). Inspection reveals that the South Wales dairy industry is obviously the more stable of the two. Each element

in the diagonal of the South Wales matrix is larger than the corresponding
one in that of the South East. The difference is quite considerable. In
each case, except for the two extreme classes, it is in excess of 0.1.
Similarly dispersion of the off-diagonal probabilities is greater in the
matrix for the South East (Table 4) than the one for South Wales (Table 1).
This can be seen from the fact that there are fewer nonzero probabilities
in the South Wales matrix, and they are arranged close to the diagonal.
Thus, according to the transition probabilities matrix for South Wales,
whatever movement takes place is assumed to be mainly between neighbouring
classes. In the South Eastern region, however, producers move between
classes which differ more widely in size.

The transition probabilities matrices for the 11 MMB regions are
presented at the end of Appendix I, in Table 12.

It should be noticed that the transition probabilities matrices all
contain nonzero entries in row 0 and column 0. Each entry in row 0 is the
assumed probability of a new producer beginning production at at a particular
level of output. Each entry in column 0 is the assumed probability of a
producer leaving production from a particular level of output. It will be
noticed that, in both row 0 and column 0 of some regional matrices, there
are probabilities associated with quite large levels of output. In row 0
of the South Wales matrix (Table 1), for example, there is a probability (.0002)
corresponding to the contingency of a producer starting from no production and
entering the industry at the "Substantial" level. In column 0, there is a
probability (.0114) of a producer ceasing production after being in the "Large" class.

This apparent suddenness with which large producers are assumed to enter
or leave the industry is to some extent an exaggeration. In the

survey, the units in which the MMB enumerates producers are farmers, partnerships and companies and not the actual farms themselves. Thus a producer is defined in terms of ownership of assets and not the assets themselves. For this reason, in the event of a dairy farm changing hands, the transaction does not appear in the MMB records as a simple transfer of ownership. It is recorded twice. It is recorded that the original producer ceases production, and at the same time the new owner suddenly enters production at the same level of output. This double counting makes no difference to estimates of output or producer numbers in any sample year, however it does distort the transition probabilities matrices by indicating a greater proportion of entrants and leavers than there actually are in terms of physical dairy farms. Hence, for predictive purposes, it exaggerates the degree of instability.

(2) The Index of Stability

The differences between the transition probabilities matrices for South Wales and the South East are pronounced. However, there are 9 other regions whose transition probabilities matrices[1] exhibit differences which are less obvious and it is hard to differentiate among them. In order to try and distinguish these differences in the degree of stability it was

---

1. presented below at the end of Appendix I

decided to use an index number. This index number consists of a weighted sum of the diagonal elements of the transition probabilities matrix. The weight used for each probability is the proportion of the producers of the region in the appropriate class in the year 1966-7[1]. The index number thus formed is to be interpreted as an index of stability defined in terms of producer numbers. The greater the stability of a regional dairy industry, the closer the index number will be to unity, and conversely, the greater the instability, the smaller it will be.

(3) <u>Regional differences of stability</u>

Regional comparisons of stability among the eleven MMB regions using the index described in the previous section showed a remarkable geographical pattern (see Table 5). Regions nearest the west coast were found to be the most stable, and those towards the South and East of the country the most unstable.

---

1. Algebraically, the expression for the index number is:

$$\text{Index} = \frac{\sum_{i=1}^{6} P_{ii} W_{i}}{\sum_{i=1}^{6} W_{i}}$$

where the $P_{ii}$ are the diagonal elements of the transition probabilities matrix $(i=1,2,3,\ldots,6)$ and $W_{i}$ is the number of producers in the $i^{th}$ class in 1966-7

TABLE 5

VALUE OF THE STABILITY INDEX FOR EACH REGION

| Region | Index |
|---|---|
| South Wales | 0.7886 |
| North Western | 0.7792 |
| North Wales | 0.7790 |
| Mid Western | 0.7623 |
| Far Western | 0.7566 |
| Eastern | 0.7418 |
| Northern | 0.7296 |
| West Midlands | 0.7136 |
| Southern | 0.7112 |
| East Midlands | 0.7108 |
| South Eastern | 0.6991 |

The greatest degree of stability was shown by South Wales, the North-west, and North Wales, where the index was claculated to be about 0.78. Next were found to be the Mid West and Far West with 0.76, and the Eastern and Northern regions with an index number of about 0.73. The West Midlands, Southern and East Midlands regions were next in the ranking , with an index valued at around 0.71. The least stable region was the South East with 0.70.

There is a noticeable similarity between this ranking, according to stability, and the ranking of regions according to the percentage of standard output deriving from the sale of milk. This means that the greater the proportion of standard output from milk in a region, the higher the index of stability.

This is of course, what might have been expected. The greater the dependence of a dairy farmer on milk production the less likely he is to change the scale of his output. Producers are heavily dependent on milk production in the regions nearest the Irish Sea and this may to a large extent explain the stability noted in these regions.

5) The Test of Constancy of the Transition Probabilities Matrices

a) The Test

Before the Markov chain technique can be used in prediction it is necessary to make the assumption that the transition probabilities remain constant over the period to which the predictions refer. For the present study it was necessary to assume that the calculated transition probabilities, derived from the sample taken between 1963 and 1967, would remain unchanged until 1975. To assess the validity of this assumption (and hence the validity of the use of Markov chains) it is necessary to demonstrate constancy over the sample period. Hence the statistical test known as chi-square was carried out for each region for the years 1963/64-1966/7.

For each region a value of the test statistic was claculated and compared with statistical tables. From this comparison the constancy or otherwise of the transition probabilities was inferred.

The principle underlying the test is that the value of the test statistic (incorporating the annual deviations of calculated probabilities from their means) reflects year-to-year changes in the estimates of the transition probabilities. If this value is very large (relative to standard chi-square tables)

See Appendix II

it can be inferred that the year-to-year variations in the transition probabilities are such as to render the assumption of constancy unrealistic.

The results of the chi-square test are presented in Table 6 below. Besides the calculated value of chi-square and its degrees of freedom, the table contains the approximate probability of the calculated value being exceeded according to the standard tabulated chi-square distribution. If this probability is small it implies a relatively high value of chi-square and therefore is evidence of significant variability in the elements of the transition probabilities matrix from year to year. Hence the assumption of a stable transition probabilities matrix is untenable and the use of Markov chains in such circumstances is unjustified.

A more technical exposition of the nature of the chi-square test in the present context is contained in Appendix II.

(b) The results of the test

In the main, the results presented in Table 6 lend limited support for the applicability of Markov chain analysis to the data collected by the Permanent Producer Survey.

In 8 regions, the chi-square values would clearly seem to support the assumption of constancy in the transition probabilities. In these regions the probability of the test statistic being exceeded is over 70 per cent. In some regions it is as high as 98 per cent or 99 per cent.

## TABLE 6

### RESULTS OF THE CHI-SQUARE TEST

| Region | Degrees of Freedom | Chi-Square Values | Approximate percentage probability of the calculated chi-square being exceeded |
|---|---|---|---|
| Northern | 44 | 39.4 | 70 |
| Northwestern | 51 | 61.9 | 20 |
| Eastern | 39 | 40.0 | 50 |
| East Midlands | 48 | 23.2 | 99 |
| West Midlands | 48 | 39.0 | 80 |
| North Wales | 33 | 18.1 | 98 |
| South Wales | 36 | 13.7 | 99 |
| Southern | 45 | 32.9 | 90 |
| Midwestern | 48 | 59.2 | 10 |
| Far Western | 36 | 17.0 | 99 |
| South Eastern | 48 | 33.8 | 95 |

In two regions the results of the chi-square test are such as to throw doubt on the appropriateness of the use of Markov chains. These are the Northwestern and Midwestern regions. For these regions the probabilities of the true chi-square value lying to the right of the calculated chi-square values are only 20% and 10% respectively.

This result for the North West is roughly consistent with that found in the earlier study[1] of the same region where the probability lying to the right of the chi-square value was found to be approximately 25% of the distribution. In view of the lack of improvement in the chi-square result, with the new data, the suggestion made in the earlier paper that the North West was likely to be the most stable region, and hence the most suited to

---

1. D. R. Colman op. cit

Markov Chain Analysis, appears to be unjustified.

It is of interest to note, without offering any explanatory hypothesis, that these two regions, whose transition probabilities are the least constant, are also the largest. In 1966-7, the Northwest accounted for 21.2% and the Mid west 10% of active producers (Table 7). However, there does not seem to be a relationship in the other nine regions between stability of transition probabilities and number of producers.

The use of Markov chains in these circumstances can be justified by reference to the fact that the regions with high values of chi-square are in the minority (3 out of 11) and represent only about one third of producers. It must, however, be borne in mind that the use of Markov chains in these regions is liable to lead to relatively larger prediction errors than in the other regions.

Results of the Prediction Procedure

(1) In terms of producer numbers the general picture which emerges is one of rapid decline both nationally and regionally. This decline may be expected to continue unremittingly (see Tables 7 and 8). In England and Wales the predicted fall in the number of producers is:

16.4 per cent. between 1966/7 and 1970/71

30.5 per cent. between 1966/7 and 1975/76

Expressing this in terms of producer numbers, there is an expected decline from:

86000 in 1966/7 to

72000 in 1970/71 and

60000 in 1975/76

## TABLE 7

### PREDICTED REGIONAL DISTRIBUTION OF PRODUCER NUMBERS

| Region | 1966-7 | | 1070-1 | | 1075-6 | |
|---|---|---|---|---|---|---|
| | Number | Percentage of National Total | Number | Percentage of National Total | Number | Percentage of National Total |
| Northern | 9874 | 11.4 | 7716 | 10.6 | 5914 | 9.8 |
| Northwestern | 18351 | 21.2 | 16246 | 22.4 | 14206 | 23.6 |
| Eastern | 3165 | 3.6 | 2489 | 3.4 | 1931 | 3.2 |
| East Midlands | 4036 | 4.7 | 3232 | 4.5 | 2501 | 4.1 |
| West Midlands | 7516 | 8.7 | 6287 | 8.7 | 5309 | 8.8 |
| North Wales | 6386 | 7.4 | 5232 | 7.2 | 4300 | 7.1 |
| South Wales | 9922 | 11.4 | 8445 | 11.6 | 7124 | 11.8 |
| Southern | 3550 | 4.1 | 2754 | 3.8 | 2154 | 3.6 |
| Mid Western | 8669 | 10.0 | 7202 | 9.9 | 6043 | 10.0 |
| Far Western | 11693 | 13.5 | 9973 | 13.8 | 8435 | 14.0 |
| South Eastern | 3577 | 4.1 | 2932 | 4.0 | 2392 | 4.0 |
| England/Wales | 86739 | 100.0 | 72508 | 100.0 | 60309 | 100.0 |

## TABLE 8

### PERCENTAGE DECLINE IN THE NUMBER OF PRODUCERS FROM 1966-7

| Region | 1970-1 | 1975-6 |
|---|---|---|
| Northern | 21.9 | 40.1 |
| Northwestern | 11.5 | 22.6 |
| Eastern | 21.4 | 39.0 |
| East Midlands | 19.9 | 38.0 |
| West Midlands | 16.4 | 29.4 |
| North Wales | 18.1 | 32.7 |
| South Wales | 14.9 | 28.2 |
| Southern | 22.4 | 39.3 |
| Mid Western | 16.9 | 30.3 |
| Far Western | 14.7 | 27.9 |
| South Eastern | 18.0 | 33.1 |
| England/Wales | 16.4 | 30.5 |

(2) So far as particular regions are concerned, the biggest rates of

decline in numbers are in the Northern, Eastern, East Midlands and Southern

regions, with a decline of : over 20 per cent by 1970-1 and

over 38 per cent by 1975-6

The smallest rate of decline is in the Northwestern region with

11.5 per cent by 1970-1 and

22.6 per cent by 1975-6

It will be seen that, even here, in the region with the slowest rate of decline, the number of producers falls by one fifth in nine years.

The remaining six regions all have intermediate rates of decline ranging between:

> 14.7 per cent and 18.1 per cent by 1970-1 and between
>
> 27.9 per cent and 33.1 per cent by 1975-6

In other words these six regions are expected to suffer a fall in numbers of about one third.

A comment which should perhaps be made here is that is is possible that the considerable decline in predicted producer numbers in all regions is to some extent an overestimate. It has been found by the MMB that the method used to select the sample has probably led to an underestimate of the number of new producers entering production each year. This being the case, it will have made a difference to the values of the transition probabilities matrices and led to an underestimate of the future number of producers in each class. However, this is a minor point because the number of producers affected by the error is likely to be very small.

(3) Besides an overall decrease in numbers, the results of the Markov chains analysis indicate changes in the regional concentration of dairy production. The number of producers may be expected to show a relative increase in four regions. This is seen in Table 7.

The number of producers expressed as a proportion of the national total increases in: the Northwestern, West Midlands, South Wales and Far Western regions, remains constant in the Mid Western region and falls in every one of the others.

This pattern would seem to lend support to the view that dairy production is becoming increasingly concentrated in those parts of the country where natural conditions are most favourable.  These natural conditions relate mainly to climate, dairy farming tending to predominate in those areas which are least suitable for arable farming; i.e. the western counties with the wetter soils and smaller average farm size than the east.

(4) Output

According to the results of the analysis, output of milk in England and Wales may be expected to rise despite the fall in producer numbers. Predicted output is presented in Table 9 and its rate of change and that of its regional components in Table 10.

It will be seen from these that the percentage change in output predicted is:

4 per cent between 1966-7 and 1970-71 and

7.3 per cent between 1966-7 and 1975-6

In 1966-7 recorded sales though the MMB were:

2040.3 million gallons and this may be expected to rise to

2162.1 million gallons in 1975-6.

It is difficult to decide whether this prediction is likely to prove realistic or not.  The only data available from the post-sample period relate to 1967-8.  In that year recorded output was 2141.1 million gallons a figure almost equal to our 1975-6 estimate.  This may indicate that the predicted rate of growth is an underestimate or that production in 1967-8 was unusually high.

## TABLE 9

### PREDICTED OUTPUT OF MILK

| Region | 1966-7 Output m.g. | 1966-7 Percentage of National Total | 1970-1 Output m.g. | 1970-1 Percentage of National Total | 1975-6 Output m.g. | 1975-6 Percentage of National Total |
|---|---|---|---|---|---|---|
| Northern | 200.2 | 9.8 | 184.2 | 8.8 | 172.1 | 8.0 |
| Northwestern | 451.9 | 22.1 | 455.9 | 21.8 | 466.3 | 21.6 |
| Eastern | 91.2 | 4.5 | 85.6 | 4.1 | 81.7 | 3.8 |
| East Midlands | 110.6 | 5.4 | 105.8 | 5.0 | 99.6 | 4.6 |
| West Midlands | 203.1 | 10.0 | 215.3 | 10.3 | 225.6 | 10.4 |
| North Wales | 88.1 | 4.3 | 93.7 | 4.5 | 101.1 | 4.7 |
| South Wales | 148.7 | 7.3 | 153.7 | 7.3 | 169.5 | 7.8 |
| Southern | 119.2 | 5.8 | 129.4 | 6.2 | 137.1 | 6.3 |
| Mid Western | 288.3 | 14.1 | 319.6 | 15.3 | 343.3 | 15.9 |
| Far Western | 197.4 | 9.7 | 195.6 | 9.3 | 206.5 | 9.6 |
| South Eastern | 141.7 | 6.9 | 156.7 | 7.5 | 159.3 | 7.4 |
|  | 2040.3 | 100.0 | 2095.5 | 100.0 | 2162.1 | 100.0 |

m.g. = millions of gallons

## TABLE 10

### PERCENTAGE CHANGE IN OUTPUT FROM 1966-7

| Region | 1970-1 | 1975-6 |
|---|---|---|
| Northern | -5.6 | -11.8 |
| Northwestern | +3.1 | +5.5 |
| Eastern | -3.9 | -8.3 |
| East Midlands | -5.0 | -10.6 |
| West Midlands | +6.8 | +11.9 |
| North Wales | +7.5 | +15.9 |
| South Wales | +7.7 | +18.8 |
| Southern | +9.8 | +16.3 |
| Mid Western | +9.8 | +18.0 |
| Far Western | +6.4 | +12.4 |
| South Eastern | +2.9 | +4.6 |
| England/Wales | +4.0 | +7.3 |

Figures produced by the Economic Development Committee for Agriculture[1]
envisage an increase for the whole of the U.K. of 15.4 per cent in the six
year period 1966-7 to 1972-3, which would be a much faster rate of growth

---

"Agriculture's Import Saving Role" Economic Development Committee for
Agriculture June 1968

than that predicted in the present study. However, it is likely that the EDCA estimates are optimistic. The publication from which they are taken is something in the nature of a policy document intended to indicate future trends in agriculture on the assumption that all opportunities for import saving are exploited and that all necessary incentives are given. The present study, however, makes the implicit assumption that no changes in the economic environment in the dairy industry will occur. When this is taken into account, it is possible that the present study, in indicating a slow rate of growth of output, shows the need for changes in government policy towards the dairy industry if it is desired that milk output should achieve a faster rate of growth.

(5) Concomitant with the rise in output nationally a movement towards regional concentration is indicated (Tables 9 and 10).

Three regions, the Northern, Eastern and East Midlands may be expected to suffer an absolute fall in output, while regions nearer the south and west show an increase far above the average for England and Wales. The regions expected to enjoy the highest rates of increase are North and South Wales, Southern and Midwestern, with:

over 7% by 1970-1

and 16% by 1975-6

(6) The pattern of regional concentration of output described in the above section, (5) is also reflected in the predicted proportional distribution of output contained in Table 9.

The four regions extending over the northern and eastern counties

(i.e. the Northern, Northwestern, Eastern and East Midlands) may be expected to experience a declining percentage of output. The share of output coming from the Southeastern region is predicted to increase between 1966-7 and 1970-1 but declines slightly thereafter.

Again, as in the case of producer numbers, the results indicate that the regional structure of output may be expected to undergo a pattern of change in which production becomes more concentrated. The counties of Wales, and the south and west of England may be expected to make a relatively larger contribution to output in the future than in 1966-7, while the northern and eastern regions show a decline.

(7) It is of interest to note that the changes which may be expected in the regional structure of output do not coincide with predicted changes in the regional distribution of producer numbers.

It will be seen that some regions increase their relative contribution to output despite a relative fall in producer numbers. This occurs in the regions in the south and west; particularly the Southeastern, Midwestern, Far Western and North Wales. In these regions it would appear that there is a stronger tendency towards an increasing size of units then elsewhere. Production becomes increasingly concentrated into the hands of a smaller number of producers whose average output is larger.

# TABLE 11

## SIZE DISTRIBUTION OF PRODUCERS

| Class | 1966-7 Number | Percentage of Total | 1970-1 Number | Percentage of Total | 1975-6 Number | Percentage of Total |
|---|---|---|---|---|---|---|
| Very Small | 16720 | 19.3 | 13474 | 18.6 | 10218 | 16.9 |
| Small | 23637 | 27.2 | 16730 | 23.1 | 12259 | 20.3 |
| Average | 19704 | 22.7 | 15430 | 21.3 | 11805 | 19.6 |
| Total | 60061 | 69.3 | 45634 | 63.0 | 34282 | 56.8 |
| Substantial | 19655 | 22.6 | 17225 | 23.8 | 14399 | 23.9 |
| Large | 5976 | 6.9 | 6919 | 9.5 | 7284 | 12.1 |
| Very Large | 1047 | 1.2 | 2694 | 3.7 | 4344 | 7.2 |
| Total | 26678 | 30.7 | 26838 | 37.0 | 26027 | 43.2 |
| Total England and Wales | 86739 | 100.0 | 72472 | 100.0 | 60309 | 100.0 |

## TABLE 12

## DISTRIBUTION OF OUTPUT BY SIZE CLASSES

| Class | 1966-7 | | 1970-1 | | 1975-6 | |
|---|---|---|---|---|---|---|
| | Output m.g. | Percentage of Total | Output m.g. | Percentage of Total | Output m.g. | Percentage of Total |
| Very Small | 71.7 | 3.6 | 57.7 | 2.8 | 43.8 | 2.0 |
| Small | 276.6 | 13.7 | 196.7 | 9.3 | 144.0 | 6.6 |
| Average | 383.8 | 19.1 | 300.2 | 14.3 | 229.2 | 10.6 |
| Total | 732.1 | 36.4 | 554.6 | 26.4 | 417.0 | 19.2 |
| Substantial | 678.1 | 33.7 | 593.4 | 28.3 | 496.4 | 23.0 |
| Large | 426.6 | 21.2 | 491.7 | 23.5 | 516.2 | 23.9 |
| Very Large | 177.4 | 8.8 | 455.8 | 21.8 | 732.4 | 33.9 |
| Total | 1282.1 | 63.6 | 1540.9 | 73.6 | 1745.0 | 80.8 |
| Total: England and Wales | 2014.2 | 100.0 | 2095.5 | 100.0 | 2162.0 | 100.0 |

m.g. = millions gallons

This concentration will be seen in Tables 11 and 12 which contain the predicted size distribution of producer numbers and output for the whole of England and Wales. The figures show quite dramatically the increasing contribution to output of the larger producers.

In terms of the number of producers (Table 11), in 1966-7, 30.7 per cent are in the largest three classes. The corresponding figure for:

1970-1 is 37.0 per cent and

1975-6 is 43.2 per cent

In terms of output (Table 12), the pattern is even more noticeable. The proportion of output coming from producers in the largest three classes in 1966-7 is 63.6 per cent

The 1970-1 figure is 73.6 per cent

and that for 1975-6 is 80.8 per cent

## CONCLUSION

The first objective of the current study was to analyse the pattern of structural change in the dairy industry and to make some forecast of its future structure on the basis of this pattern. The principal conclusions in this respect were, firstly, at the national level, a small rate of growth of output accompanied by a fairly massive rate of decline in the number of active producers. At the micro level the analysis forecasts a considerable movement towards industrial concentration. A rapidly increasing proportion of output may be expected from larger producers. Regionally, a slight shift in the distribution of production in favour of the regions lying in the south and west of the country at the expense of those in the east was indicated.

These results all depend on the underlying assumption that the economic conditions which prevailed between 1963 and 1967 will remain unchanged over the period of prediction. The small rate of growth of output predicted on this assumption might be regarded from the policy point of view as being inadequate. This being so, some change in government policy towards the dairy industry would be needed to achieve a more satisfactory rate of growth of milk output.

The second objective of the study was concerned with the methodology employed. This was to assess the relevance of the assumption of a constant, and discernable pattern of change in industrial structure, upon which the applicability of Markov chain analysis relies. The validity of this assumption was tested by applying the statistical chi-square test to the sample data in each region. The information thus supplied was necessarily limited since

it was subject to the reservations which surround all significance tests.
Within these restrictions, however, little evidence was found which
would lead to a rejection of the use of Markov chain analysis. Only in
two regions did the calculated value of chi-square approach a value which
would be critical at the usual levels of significance. In these, the North-
western and Midwestern regions, the results of the chi-square test were
such as to cast some doubt on the validy of prediction by Markov chains.

For this reason it should, perhaps, be borne in mind that in these regions
prediction is likely to be subject to a greater degree of error than in the
regions where the chi-square results were more satisfactory. These errors
may well have a strong effect on the prediction of aggregate output and
producer numbers.

A valid criticism which may be levelled at the general line of approach
adopted is that the variable by which producers were classified was annual
output, and not herd size. Criticism of the output classification can be
made on the grounds that annual milk output is not a decision variable wholly
within the control of the producer. Output is open to influences outside
the control of the producer - in particular the weather - to which herd size
is not susceptible. For this reason prediction based on a classification by
output would be likely to lead to greater inaccuracy than prediction on the
basis of herd size. The classification of producers by output was used in
this study of necessity due to the nature of the data, since, although herd
size data do exist, the samples from which they are obtained are less
representative than the Permanent Producer Survey.

A Further drawback is in a theoretical question attaching to the application

of Markov chain analysis. This is the problem of taking account of the
influence of new entrants to the industry. There is no really satisfactory
theoretical solution to it, but it is hoped that the expedient employed here
(described in Appendix I(2)) adequately takes account of this effect for the
purpose of prediction.

## APPENDIX I

## METHOD OF PREDICTION

### (1) Deviation of Transition Probabilities

An algebraic treatment of the derivation of a transition probabilities matrix is presented here in an effort to clarify the intuitive treatment in the text.

Firstly, sample producers were assigned to classes on the basis of recorded output in the year t-1. The allocation was repeated for output in the following year, t. A matrix was then drawn up each of whose rows , i, represented a class in year t-1, and each of whose columns represented a class in year t. There are seven classes, one of which denotes zero output, hence $= i,j = 0,1,2, \ldots,6$. Each of the elements of the matrix is the number of sample producers, $N_{ijt}$, following a similar career in moving from class i in year t-1 to class j in year t (t=1,2,3, etc.)

From this matrix, $\| N_{ijt} \|$, was calculated the transition probabilities matrix for the year t, denoted by $\| P_{ijt} \|$

$$P_{ijt} = \frac{N_{ijt}}{\sum_{i=0}^{6} N_{ijt}} \qquad \text{all } i,j,t$$

Each element of the matrix thus obtained is the probability of a producer ending in class j at time t; conditionally upon being in state i at time t-1. Each row is a conditional probability distribution thus:

$$0 < P_{ijt} < 1 \qquad \text{for all } i,j$$

and the sum of the $P_{ijt}$ equals one.

$$\sum_{j=0}^{6} P_{ijt} = \sum_{j=0}^{6} \frac{N_{ijt}}{\sum_{j=0}^{6} N_{ijt}} = \frac{\sum_{j=0}^{6} N_{ijt}}{\sum_{j=0}^{6} N_{ijt}} = 1$$

all i,t

The elements of the transition probabilities matrix used in prediction are averages taken over the whole sample period, calculated from the elements of the $\left\| N_{ijt} \right\|$ matrix.

The $\left\| N_{ij} \right\|$ matrix was obtained by aggregating the $\left\| N_{ijt} \right\|$ matrix over the sample period, from the year t=1 to t=T.

In this oase T=3

$$N_{ij} = \sum_{t=1}^{T} N_{ijt} \qquad \text{for all i,j}$$

The predictive transition probabilities matrix, $\left\| P_{ij} \right\|$ was derived from $\left\| N_{ij} \right\|$ by means of the definition:

$$P_{ij} = \frac{N_{ij}}{\sum_{j=0}^{6} N_{ij}} \qquad i,j = 0,1, \ldots,6$$

It will be seen that each element of the matrix $\left\| P_{ij} \right\|$ is a probability; i.e. $0 \ P_{ij} \ 1 \quad i,j = 0,1, \ldots,6.$

and that the matrix has the desired property that each row is a conditional probability distribution, since clearly

$$\sum_{j=0}^{6} P_{ij} = 1 \qquad i = 1, \ldots,6;$$

## (2) The Entrants Problem

A serious problem in the application of Markov chain analysis attaches
to the value of the probability to be assigned to the cell at the intersection
of row 0 and column 0 of the transition probabilities matrix. This value is
the probability that a producer who is a potential entrant to the dairy industry
at the beginning of a year will _not_ have entered the industry by the end of the
year, i.e. remains a potential entrant. The words "assigned to the cell"
are used advisedly since it is difficult to define a potential entrant. It
is difficult to conceive of taking a sample of potential entrants in the same
way as one could of actual entrants. Indeed no such sample group has been
considered in collecting the data used in this study. As a consequence of
this, sample data were available for all cells in the $\| N_{ij} \|$ matrix except
$N_{00}$, without which it is not possible to estimate the probabilities $P_{0j}$

since these depend upon the row total $\sum_{j=0}^{6} N_{0j}$ the number of potential entrants

at the beginning of the year). Thus it has been necessary to select an arbitrary
value for the number of potential entrants in order that the first row
probabilities may be derived.

The arbitrary value selected for the number of potential entrants turns
out to be of considerable importance. For, as Stanton and Kettunen[1] have

---

1. B. F. Stanton and L. Kettunen  "Potential Entrants and Projections in Markov
Process Analysis" Journal of Farm Economics  Vol. 49, pp. 633-643, August 1967

demonstrated the number chosen materially influences the predictions of the

path followed by the industry. They advise that a large number of potential

entrants be assumed for industries which are perfectly competitive, as this

complies with the theoretical economic conditions for this form of industrial

structure. Such a choice therefore seems to be indicated for the dairy

industry which has a highly atomistic structure.

Two further considerations lead to the choice of a large number of

potential entrants. The first of these is that if a low number were chosen

the probability of entry would become high, since a large proportion of the

potential entrants would in fact be observed to enter the industry in a

given year. Thus if there were few farms leaving the industry the pool of

entrants would quickly reduce to an unreasonably low level. Choice of a

large pool of entrants with a resultant low probability of entry would

result in a steady flow of entrants (such as might be expected) which would

in the short-run be only marginally affected by the rate of outflow of farms.

The second consideration is that choosing a high number of potential entrants

leads to a low probability of immediate re-entry by farms which have just

left the industry. (It does seem reasonable to assume that many producers who

leave the industry are unlikely to re-enter immediately). Effectively a

higher than first order re-entry condition is imposed upon producers who

have left the industry by the choice of a large pool of entrants - that is to

say that the probability of their re-entry is forced down to a very low level

by this device.

A brief example may suffice to explain this. Assume that 20 producers

are observed in the sample data to enter the smallest class ("Very Small")

and that no producers enter any other classes. Hence $N_{oi} = 20$ and $N_{oj} = 0$

for $j = 2, 3, \ldots, 6$. If we assume that the pool of entrants at the beginning of

the period was 40, then since 20 are observed moving into the "Very Small"

class, 20 must have remained in class 0. Thus $P_{00} = 0.5 \left( \text{ie.} \sum_{j=0}^{6} \frac{N_{00}}{N_{oj}} = \frac{20}{40} \right)$

and $P_{oi} = 0.5$. If we raise the assumed pool of entrants to 4,000 then

$P_{00} = 0.995$ and $P_{oi} = 0.005$. In other words the probability that any farmer

who is a potential milk producer will commence production in the current

period has been reduced arbitrarily from 0.5 to 0.005. It should be

remarked, however, that the <u>actual number of entrants</u> during the initial

period is the same in both cases.

The figure chosen for the size of the pool of potential entrants in

1963 was approximately the number of active producers in each region

<u>multiplied by three</u> - approximate in that the number was rounded to the

nearest ten thousand. Thus a region with an active producer population of

approximately 10,000 was assumed to have a pool of potential entrants of

30,000. In estimating the sample probabilities this number was reduced in

the ratio of the sample number of entrants to the population number of

entrants.

The number assigned to the pool in the start vector (used as the basis of

prediction) was the estimated number of potential entrants in 1966. This figure

was obtained from the assumed 1963 pool by making allowance in each year of

the sample for the net increase or decrease which resulted from producers leaving

and entering production.

It should be noted that fairly large differences in the assumed number of potential entrants have little influence on the predicted size distribution in the first few periods, although major differences occur for later periods. As this study was only concerned with relatively short period prediction this problem of divergence was not a serious one.

(3) The Prediction procedure

Prediction of the structure of the industry was carried out using the transition probabilities matrix, $\left\| P_{ij} \right\|$, and a start vector, $W_o$, representing the distribution of producers in the base year of prediction. $W_o$ is a row vector of seven elements.

$$W_o = ( W_{oo}, W_{o1}, W_{o2}, \ldots, W_{o6})$$

Each element is the number of producers in the region contained in each class. Thus, for instance, $W_{o1}$ is the number in the "Very Small" class. $W_{oo}$ represents the hypothetical number of producers with zero production. (Its derivation is discussed in section (2) above).

A description of the iterative prediction procedure involves the use of matrix algebra.

Denote the matrix $\left\| P_{ij} \right\|$ by $P$. Then the size distribution of producers in year 1 (i.e. the first year of prediction) is given by the relation:

$$W_1 = W_o P$$

i.e. 
$$(W_{1o}, W_{11}, \ldots, W_{16}) = (W_{oo}, W_{o1}, \ldots, W_{o6}) \begin{Vmatrix} P_{oo} & P_{o1} & \ldots\ldots & P_{o6} \\ P_{1o} & P_{11} & \ldots\ldots & P_{16} \\ & & & \\ P_{6o} & P_{61} & \ldots\ldots & P_{66} \end{Vmatrix}$$

Similarly $W_2 = W_1 P$

$$= W_o PP$$

$$= W_o P^2$$

In general:

$$W_t = W_o P^t$$

The major assumption underlying this procedure is that the elements of P (i.e. the individual) $P_{ij}$'s represent the true transition probabilities and that these remain constant throughout the sample and prediction periods.

This assumption may, of course, be false, because the pattern of change is not regular and the transition probabilities change over time. The extent to which the assumption is justified in a particular case can be guaged by use of the chi-square test described in Appendix II.

## Appendix II

### The Chi-Square Test

(1) The test

The rather weak chi-square test[*] used in the current study is applied

to the null hypothesis =

$$H_o : P_{ijt} = P_{ij} \qquad i,j = 0,1,2, \ldots,6$$
$$t = 1,2,3$$

with the alternative hypothesis :

$$H_1 : P_{ijt} \neq P_{ij}$$

The test statistic is chi-square:

$$X^2 = \sum_{i=0}^{6} \sum_{j=0}^{6} \sum_{t=0}^{3} \frac{(P_{ijt} - P_{ij})^2}{P_{ij}} \sum_{j=0}^{6} N_{ijt}$$

with 7x6x3 degrees of freedom.

The test is weaker than its usual application to goodness of fit, because

in the current study the hypothesis which it is hoped will be supported by

the result of the test, is the null hypothesis. This differs from the

more usual approach to significance testing where inferences are only drawn

when the null hypothesis is rejected in favour of the alternative. The

weakness in the present test arises because it is not possible to show that

the null hypothesis is accepted. If the value of the test statistic is outside

the critical region this does not lead to acceptance of the null hypothesis.

[*] The test is one of two applications to Markov chains devised by T. W. Anderson
and L.A. Goodman, "Statistical Inference About Markov Chains", The Annals of
Mathematical Statistics, March 1957. The other is the likelihood ratio test,
but is inapplicable here because some of the expected frequencies are zero.

It is consistent with either the null or the alternative hypothesis being true and the only conclusion which can be reached is that the null hypothesis is not rejected.

(2) <u>The results</u> When viewed in this light all that can be said about the values obtained for the test statistic in the current study is that they do not lead to rejection of the null hypothesis at the 5% level. In the Midwestern region, however, the null hypothesis is rejected at the 10% level, and in the Northwestern region at the 20% level. In all the other regions the corresponding level of significance is at least 50% and in some regions it is over 95%.

It can thus be said of these results, if the usual 5% level of significance is chosen, that they are not inconsistent with the null hypothesis (that the transition probabilities are constant over time). The results do not give support to the conclusion that the null hypothesis is false and therefore do not invalidate the use of Markov Chains.

(3) <u>The modified statistic</u> The values of chi-square presented in Table 6 were calculated using a modification of the expression given above. The reason for making the modification was the presence of a large number of very small theoretical frequencies in the matrix $\left\| N_{ij} \right\|$ When such a situation occurs the use of the unmodified formula exaggerates the value of chi-square in an upwards direction.

The modification consists, simply, of amalgamating certain elements of the $\left\| N_{ij} \right\|$ (and $\left\| N_{ijt} \right\|$) matrix in such a way as to fulfil more closely

the conditions laid down by Cochran[*], and then to recompute the elements of the $\|P_{ij}\|$ (and $\|P_{ijt}\|$ ) matrix. From these modified values the test statistic is calculated and the test carried out.

* W. G. Cochran, "Some Methods for Strengthening the Common chi-square Tests" Biometrics, Vol. 10. 1954. He suggests that no more than about 20% of theoretical frequencies should be less than 5, and that if this condition is not met, classes should be merged.

Appendix III

REGIONAL TRANSITION PROBABILITIES MATRICES

NORTHERN REGION

TABLE 12A

| | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| Class in first Year | O | Very Small | Small | Average | Substantial | Large | Very Large |
| O | .9954 | .0024 | .0020 | 0 | .0002 | 0 | 0 |
| Very Small | .1928 | .6749 | .1295 | .0028 | 0 | 0 | 0 |
| Small | .0254 | .1350 | .7006 | .1350 | .0039 | 0 | 0 |
| Average | .0122 | .0213 | .1189 | .7043 | .1433 | 0 | 0 |
| Substantial | .0038 | .0077 | .0077 | .0923 | .8269 | .0615 | 0 |
| Large | .0073 | .0073 | 0 | 0 | .0584 | .8540 | .0730 |
| Very Large | 0 | 0 | 0 | 0 | .0097 | .0680 | .9223 |

## TABLE 12B

## NORTH WESTERN

| Class in First Year | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9964 | .0020 | .0010 | .0001 | .0001 | 0 | 0 |
| Very Small | .2021 | .6937 | .0917 | .0125 | 0 | 0 | 0 |
| Small | .0321 | .0929 | .7643 | .1071 | .0036 | 0 | 0 |
| Average | .0194 | .0291 | .1040 | .7323 | .1137 | .0014 | 0 |
| Substantial | .0015 | .0015 | .0076 | .0787 | .8593 | .0514 | 0 |
| Large | .0054 | .0054 | 0 | 0 | .0924 | .8315 | .0652 |
| Very Large | 0 | 0 | 0 | 0 | .0054 | .0649 | .9297 |

## TABLE 12C

### EASTERN

| Class in First Year | Class in Second Year | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9958 | .0019 | .0014 | .0005 | .0005 | 0 | 0 |
| Very Small | .2194 | .7226 | .0581 | 0 | 0 | 0 | 0 |
| Small | .0724 | .1316 | .6842 | .0921 | .0197 | 0 | 0 |
| Average | .0305 | .0305 | .1069 | .7252 | .1069 | 0 | 0 |
| Substantial | .0496 | .0083 | .0248 | .1074 | .7438 | .0661 | 0 |
| Large | 0 | .0115 | 0 | 0 | .0575 | .8736 | .0575 |
| Very Large | 0 | 0 | .0061 | 0 | .0061 | .0429 | .9448 |

## TABLE 12D

## EAST MIDLANDS

| Class in | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| First Year | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9948 | .0020 | .0012 | .0008 | .0008 | .0004 | 0 |
| Very Small | .3082 | .5568 | .1136 | .0114 | 0 | 0 | 0 |
| Small | .0698 | .1240 | .6822 | .1240 | 0 | 0 | 0 |
| Average | .0422 | .0422 | .1325 | .6325 | .1446 | .0060 | 0 |
| Substantial | .0292 | .0058 | .0175 | .0994 | .8070 | .0409 | 0 |
| Large | .0265 | 0 | 0 | .0177 | .0708 | .7876 | .0973 |
| Very Large | 0 | 0 | .0241 | 0 | 0 | .0602 | .9157 |

## TABLE 12E

### WEST MIDLANDS

| Class in First Year | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| | O | Very Small | Small | Average | Substantial | Large | Very Large |
| O | .9939 | .0034 | .0012 | .0010 | 0 | .0005 | 0 |
| Very Small | .2625 | .6083 | .1125 | .0083 | .0042 | .0042 | 0 |
| Small | .0576 | .1695 | .6339 | .1356 | .0034 | 0 | 0 |
| Average | .0381 | .0381 | .1176 | .6713 | .1592 | 0 | 0 |
| Substantial | .0196 | .0131 | .0098 | .0817 | .8235 | .0523 | 0 |
| Large | 0 | 0 | 0 | 0 | .0806 | .7984 | .1210 |
| Very Large | 0 | .0070 | 0 | 0 | .0070 | .0704 | .9155 |

## TABLE 12F

## NORTH WALES

| Class in First Year | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9964 | .0027 | .0005 | .0002 | .0002 | 0 | 0 |
| Very Small | .1214 | .8010 | .0698 | .0052 | .0026 | 0 | 0 |
| Small | .0283 | .1368 | .7406 | .0896 | .0047 | 0 | 0 |
| Average | .0157 | .0157 | .1024 | .7559 | .1102 | 0 | 0 |
| Substantial | .0104 | .0104 | .0208 | .0625 | .7917 | .1042 | 0 |
| Large | .0060 | 0 | 0 | 0 | .0361 | .9096 | .0482 |
| Very Large | 0 | 0 | 0 | 0 | 0 | .0625 | .9375 |

## TABLE 12G

## SOUTH WALES

| Class in First Year | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9975 | .0018 | .0003 | .0002 | .0002 | 0 | 0 |
| Very Small | .1250 | .7780 | .0896 | .0075 | 0 | 0 | 0 |
| Small | .0217 | .0924 | .8007 | .0815 | .0036 | 0 | 0 |
| Average | .0153 | .0122 | .0948 | .7645 | .1131 | 0 | 0 |
| Substantial | .0217 | 0 | .0072 | .0797 | .8116 | .0797 | 0 |
| Large | .0114 | 0 | 0 | .0227 | .0568 | .8636 | .0455 |
| Very Large | 0 | 0 | 0 | 0 | 0 | 0 | 1.0 |

## TABLE 12H

### SOUTHERN

| Class in First Year | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9964 | .0013 | .0007 | .0007 | .0003 | .0003 | .0003 |
| Very Small | .3300 | .6000 | .0600 | .0100 | 0 | 0 | 0 |
| Small | .0833 | .1742 | .5682 | .1591 | .0152 | 0 | 0 |
| Average | .0548 | .0411 | .1233 | .6507 | .1301 | 0 | 0 |
| Substantial | .0128 | .0128 | .0256 | .0705 | .8205 | .0577 | 0 |
| Large | .0147 | 0 | .0074 | .0147 | .0221 | .8088 | .1324 |
| Very Large | .0057 | 0 | 0 | 0 | 0 | .0747 | .9195 |

## TABLE 12I

## MIDWESTERN

| | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| Clas in First Year | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9965 | .0017 | .0007 | .0003 | .0006 | 0 | .0003 |
| Very Small | .2674 | .6203 | .0963 | .0160 | 0 | 0 | 0 |
| Small | .0524 | .1008 | .6976 | .0411 | .0081 | 0 | 0 |
| Average | .0299 | .0240 | .1018 | .7036 | .1407 | 0 | 0 |
| Substantial | .0122 | .0122 | .0024 | .0657 | .8224 | .0779 | .0073 |
| Large | 0 | 0 | 0 | .0134 | .0872 | .8456 | .0537 |
| Very Large | .0067 | 0 | 0 | 0 | 0 | .0505 | .9428 |

## TABLE 12J

## FAR WESTERN

| Class in First Year | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9953 | .0033 | .0012 | .0002 | 0 | 0 | 0 |
| Very Small | .1441 | .7517 | .0972 | .0052 | .0017 | 0 | 0 |
| Small | .0302 | .1147 | .7284 | .1227 | .0040 | 0 | 0 |
| Average | .0122 | .0091 | .1037 | .7591 | .1159 | 0 | 0 |
| Substantial | .0255 | 0 | .0127 | .0892 | .8153 | .0573 | 0 |
| Large | .0089 | .0089 | 0 | 0 | .0625 | .8571 | .0625 |
| Very Large | 0 | 0 | 0 | 0 | .0333 | .0667 | .9000 |

## TABLE 12K

## EAST MIDLANDS

| | Class in Second Year | | | | | | |
|---|---|---|---|---|---|---|---|
| Class in First Year | 0 | Very Small | Small | Average | Substantial | Large | Very Large |
| 0 | .9900 | .0050 | .0011 | .0017 | .0011 | .0011 | 0 |
| Very Small | .4086 | .5054 | .0430 | .0215 | .0108 | .0108 | 0 |
| Small | .0738 | .1393 | .6475 | .1230 | .0164 | 0 | 0 |
| Average | .0752 | .0451 | .1203 | .6541 | .1053 | 0 | 0 |
| Substantial | .0335 | .0239 | .0287 | .0670 | .7512 | .0957 | 0 |
| Large | .0079 | .0236 | .0079 | .0236 | .0945 | .7402 | .1024 |
| Very Large | 0 | 0 | 0 | 0 | .0045 | .0315 | .9640 |