MANCHES

76/4

Manchester Working Papers in Agricultural Economics

A Model of the U.K. Horticultural Sector

M.P. Burton & J.P. Martin

WP 86/4

£3.50

Department of Agricultural Economics

Faculty of Economic and Social Studies
University of Manchester.
Manchester U.K.

A Model of the U.K. Horticultural Sector

M.P. Burton & J.P. Martin

WP 86/4

£3.50

A MODEL OF THE UK HORTICULTURAL SECTOR

24-6-86

M.P. Burton & J.P. Martin[$]

Introduction

This working paper outlines the Horticultural model that has been estimated for use in the model of U.K. agriculture currently under development at Manchester. It consists of 5 sections:

1) An outline of the Horticultural sector in the U.K., and its relative importance.

2) A description of the Multi-Nomial Logit (MNL) model used in the land allocation model.

3) The estimated model, which uses a 33 crop classification of Horticulture. The parameter estimates are reported for the area equations, and also for the equations determining output sold and price of each of the commodities. The system is completed by a number of accounting equations that accomodate any residual elements, and which also aggregate the revenue generated at the crop level up to the Horticulture level.

4) A truncated model is presented, which uses the top levels of the full model only. This determines the area of Orchard Fruit, Soft Fruit, Vegetables and Protected Vegetables. Equations are also estimated for returns per hectare for each of these four aggregates, allowing total revenue to be determined.

5) The performance of the two models in simulating horticultural revenue is compared.

[$] The authors accept joint responsibility for the contents of this paper, and would like to acknowledge the advice given by Trevor Young on the estimation of the Multi Nomial Logit Model.

1) HORTICULTURE IN THE U.K.

The definition of Horticulture used in the Annual Review covers vegetables, fruit and non-edible crops but it excludes potatoes and hops. The diversity of crops contained in these classifications is large. For fruit, one can identify 24 different crops from the publication 'Horticultural Statistics', although a number of these are different varieties of cooking and dessert apple. At a more aggregate level, it comprises Orchard Fruit (cooking and dessert apples, pears, cider apples and perry pears, plums and cherries) and Soft Fruit (strawberries, raspberries, blackcurrants and 'others').

The vegetable sector consists of two groups: field crops and protected crops. Again, there are a large number of different crops, with some 20 grown in the open and 4 protected crops. Basic Horticulture Statistics identifies an equally wide range of non-edible crops (21 types), although revenue figures are given for aggregates (flowers in the open, flower bulbs, hardy nursery and protected crops). This brief review indicates the range of products labelled under Horticulture; from extensive field crops to those grown under glass, from the multiple cropping systems of lettuce to the perennial crops.

In terms of the 1986 Annual Review's Table 22 (Output, Input and Farming Income) Horticulture is not an inconsiderable element. Table 1 gives some of the basic data for 1984, and indicates that Horticulture generates some 11% of Final Output, and, in terms of output, is a little over 50% of the size of total cereals. The largest single element within Horticulture, Vegetables, also compares favourably with other activities, being 82% of the size of Barley, and being larger than Fat sheep and Lambs, and Poultry, and Eggs. In terms of agricultural area it is not so significant, reflecting the high returns per hectare obtained in Horticulture. Thus, in 1984, total horticulture accounted for only 1% of total agricultural area, but 11% of total output.

Table 1

OUTPUT (Revenue) for selected crops, 1984

	£m	Hort, as a %	Veg as a %
HORTICULTURE	1252	1	0.62
VEGETABLES	778	-	1.00
TOTAL CEREALS	2424	0.52	0.32
WHEAT	1447	0.86	0.54
BARLEY	947	1.32	0.82
POULTRY	674	1.86	1.15
EGGS	554	2.26	1.40
MILK	2338	0.54	0.33
FAT CATTLE	1938	0.65	0.40
FAT SHEEP & LAMBS	557	2.25	1.40
FINAL OUTPUT	11650	0.11	0.07

Source: Annual Review, 1986

AREA for selected crops 1984

	'000 ha		Hort as a %
HORTICULTURE		218	1.00
VEGETABLES	148		-
ORCHARDS	39		-
SOFT FRUIT	16		-
UNDER GLASS	2		-
NON-EDIBLE	12		-
TOTAL CEREALS		4036	0.05
WHEAT		1939	0.11
BARLEY		1978	0.11
TOTAL AREA		17501	0.01

Source: June Census, 1984

2) THE THEORETICAL MODEL

The model used to determine the areas of particular crops is Theil's Multi-nomial Logit extension of the linear logit model. The method has been successfully used by Bewley, Colman and Young (forthcoming) to allocate cereal areas, and by Bewley and Young (forthcoming) to determine meat expenditures. The following outline of the model is drawn from these works, and the interested reader is referred to those papers for a more extensive discussion of the modelling technique. The implicit assumption of the model is that the decision process is a two (or more) stage procedure, whereby a pre-determined area is allocated between a number of competing uses.

Let TA be the total area to be allocated, and A_i the area of a particular crop, then the share allocated to crop i (W_i) is given by

$$W_i = A_i/TA \qquad 1)$$

and it is hypothesised that

$$W_i = \frac{e^{f_i + u_i}}{\Sigma_{j=1}^{n} e^{f_j + u_j}} \qquad 2)$$

where n is the number of activities. The functions f_j are then specified as functions of whatever economic or other factors that may determine the allocation of area to a particular crop. The advantage of this specification is that the shares are bounded by 0 and 1, and are constrained to add up to 1, (both for estimation and simulation). The disadvantage of the method is that, if share equations are estimated directly, there are cross equation covariances in the error terms which would require an appropriate estimation procedure. In order to avoid this a transformation is undertaken.

Let $$Ln(\tilde{W}) = \frac{1}{n}.\Sigma_{j=1}^{n} Ln(W_i) \qquad 3)$$

Then. $$Ln(W_i/\tilde{W}) = f_i - \bar{f} + u_j - \bar{u} \qquad 4)$$

where $\bar{f} = \frac{1}{n}\Sigma_{j=1}^{n} f_j$ and $\bar{u} = \frac{1}{n}\Sigma_{j=1}^{n} u_j$

So, if f_i is defined as being a function of (normalised) returns per hectare, i.e.

$$f_i = a_0 + \Sigma_{j=1}^{n-1} a_j.Ln(RET_{jt-1}/RET_{nt-1}) + u_i \qquad 5)$$

the transformed model becomes

$$Ln(W_i/\tilde{W}) = \partial_0 + \Sigma_{j=1}^{n=1} \partial_j.Ln(RET_{jt-1}/RET_{nt-1}) + v_i \qquad 6)$$

where the parameters are now defined as deviations from their mean values, and v_i is independent between equations.

To this basic model one can add whatever refinements one requires. For example, weather or indices of relative costs may affect the areas planted to each crop. One option that has been utilized in the model is the possibility that the shares will vary with the total area planted. Thus, equation 6) becomes

$$\mathrm{Ln}(W_i/W^{\sim}) = \partial_0 + \Sigma_{j=1}^{n-1} \partial_j.\mathrm{Ln}(RET_{jt-1}/RET_{nt-1}) + b_i.\mathrm{Ln}\ (TA) + v_i \qquad 7)$$

The effect of this is that as the total area expands, the allocation of the area moves in the favour of a particular crop.

The other modification to the basic model that has been used is the introduction of dynamics into the specification. One method is to introduce constrained dynamics. Equation 6) would then become

$$\mathrm{Ln}(W_i/W^{\sim}) = \partial_0 + \Sigma_{j=1}^{n-1} \partial_j.\mathrm{Ln}(RET_{jt-1}/RET_{nt-1}) + g.\mathrm{Ln}(W_i/W^{\sim})_{t-1} + v_i \qquad 8)$$

This is a constrained specification, because the coefficient on the lagged dependent variable (g), has to be constrained to be equal across all equations (see Bewley, Colman and Young).

If an unconstrained specification of the dynamics is used then n-1 lagged dependents are included in each equation. (One has to be excluded in order to avoid perfect correlation between the regressors, as the sum of the n normalised shares is unity). Equation 6) then becomes

$$\mathrm{Ln}(W_i/W^{\sim}) = \partial_0 + \Sigma_{j=1}^{n-1} \partial_j.\mathrm{Ln}(RET_{jt-1}/RET_{nt-1}) + \Sigma_{j=1}^{n-1} g_j.\mathrm{Ln}(W_i/W^{\sim})_{t-1} + v_i \qquad 9)$$

This gives us six possible combinations of dynamics and explanatory variables. These can be represented as follows

	No Dynamics	Constrained Dynamics	Unconstrained Dynamics
No Total Area			
With Total Area			

3) AN APPLICATION TO THE HORTICULTURAL SECTOR

Given the large number of commodities identified within the overall grouping 'Horticulture' it is not possible to estimate the model as one unit. Instead, a recursive structure is established. Table 2 gives the crop groupings that have been used in the estimation of the model. It should be noted that some aggregation has taken place (in particular in the apple and pear groups) and that some minor crops have been excluded. The model operates in a number of stages. Thus, at the first stage, Horticultural area (area 60) is allocated between 4 alternative uses, Orchard (50), Soft Fruit (42), Vegetables (51), and Protected Vegetables (47). One can then allocate these sub-areas further, for example Orchard is split into Hard Orchard (40) and Soft Orchard (41), taking the area of Orchard Fruit as exogenous.

In this way one can move down to the crop level, giving 11 Multi-nomial Logit models. It should be noted that the non-edible sector (52) has been excluded from the analysis, as the data is not available in a form that is compatible with the other crops.

Each of the 11 models has been estimated, using each of the six specifications noted above. However, it has not been possible to aggregate all 11 models into a single model for simulation purposes, because the size of the resulting model exceeds the present limit of the program (PRODUCE) being used.

Table 2 Crop Groupings

Crop	Group	Group	Group
1 Dessert Apples	40 Hard Orchard	50 Orchard	60 HORTICULTURE
2 Cooking Apples			
3 Pears			
4 Cider Apples and Pears			
5 Plums	41 Soft Orchard		
6 Cherries			
7 Strawberries		42 Soft Fruit	
8 Raspberries			
9 Blackcurrants			
10 Others			
11 Beetroot	43 Roots	51 Vegetables	
12 Carrots			
13 Parsnips			
14 Turnips			
15 Onions, dry			
16 Onions, green			
17 Brussels	44 Brassicas		
18 Cabbage			
19 Cauliflower			
21 Broad Beans	45 Legumes		
22 Runner Beans			
23 Peas (marketed)			
24 Peas (processed)			
25 Asparagus	46 Others		
26 Celery			
27 Leeks			
28 Lettuce			
29 Rhubarb			
30 Watercress			
31 Tomatoes		47 Protected Vegetables	
32 Cucumbers			
33 Lettuce			
34 Mushrooms			
35 Flowers & Bulbs	52 Non-Edibles		
36 Nursery			
37 Protected Crops			

Efforts are being made to extend this limit to allow the full model to be run, but for the moment we have had to operate with a reduced model by excluding some of the lower levels. Thus in the discussion that follows, the "full" model refers to a system of 5 sub-models. Diagramatically this appears as:

<u>Figure 1</u> <u>The "Full" Model</u>

```
                               Horticulture (60)
          _____________________________|_____________________________
          |                    |                    |                   |
          |                    |                    |                   |
      Orchard(50)        Soft Fruit(42)      Vegetables(51)        Prot.Veg(47)
          |             /    |    |    \            |             /    |    |    \
          |          (7)  (8)  (9)  (10)            |          (31) (32) (33) (34)
      ____|____                                _____|_____
      |       |                                |   |   |   |
    (40)    (41)                            (43) (44) (45) (46)
```

The next problem is the selection of the preferred specification from the six estimated for each model. One criterion is to use a log likelihood test, but an alternative is to look at the simulation performance of the model, as it is the dynamic properties that will be important in any forecasts. The first two columns of Table 3 give the U2 statistics for the dynamic simulations for two alternative forms of the model. Note that this is a full simulation, with the areas generated at the first level feeding down to the second. The "Max. L.L." form uses the best logit model based on the log likelihood test, and the specification used is shown at the foot of the table. Although these results look quite acceptable (given that returns are being held exogenous) the model has some undesirable properties. It was found, for some lower level sub-models, that by relaxing some of the restrictions that were accepted by the log likelihood tests the simulation performance (as measured by the U(2) statistics) improved. Moreover, it was also discovered that the top level model was dynamically unstable (i.e. if returns were held constant at their 1982 levels, all of the horticultural area was allocated to 'soft fruit' by the year 2000). As this behaviour was thought to be unsatisfactory, additional specifications of the top level model were tried. The selection criterion adopted was lexiographic, based on long run stability, and then minimization of the within period U(2) statistic. The 'best' top level model found used constrained dynamics, and an additional normalised returns variable (that of the Orchard fruit), lagged two periods. This latter variable was chosen because of the possible need to allow a different adjustment path in the Orchard sector. This model is termed the 'stable form' model and the U(2) values associated with it are also given in Table 3 below.

Table 3 U2 Statistics, Dynamic Simulations 1965 to 1982

AREA, EXOGENOUS RETURNS

	Max L.L. Form	Stable Form	Final Form
HORTICULTURE			
Orchard	0.8327	1.3807	0.4776
Soft Fruit	0.2735	0.2810	0.3501
Vegetables	0.2019	0.2949	0.0291
Prot. Veg.	0.5735	0.6741	0.6447
Orchard			
Hard Orchard	0.8984	1.4548	0.5746
Soft Orchard	0.6144	0.9744	0.3199
Soft Fruit			
Strawberries	0.5513	0.5228	0.6150
Raspberries	0.6679	0.6217	0.5056
Blackcurrants	0.6692	0.8602	0.6707
Others	0.7889	0.6283	0.7660
Vegetables			
Roots	0.4333	0.5010	0.4087
Brassicas	1.0799	0.9158	0.7967
Legumes	0.6255	0.6251	0.6233
Others	0.8868	0.9146	0.8707
Protected Veg.			
Tomatoes	0.8143	0.7044	0.7849
Cucumbers	0.5969	0.6848	0.6609
Lettuce	0.5468	0.6357	0.6426
Mushrooms	0.6826	0.6512	0.7571

Maximum Log Likelihood Model

HORTICULTURE	unconstrained dynamics, with area
ORCHARD	unconstrained dynamics, without area
SOFT FRUIT	unconstrained dynamics, without area
VEGETABLES	constrained dynamics, without area
PROT. VEG.	unconstrained dynamics, with area

Table 3 cont.

Stable Model

HORTICULTURE	constrained dynamics, with area and an additional lagged return variable
ORCHARD	unconstrained dynamics, without area
SOFT FRUIT	unconstrained dynamics, with area
VEGETABLES	unconstrained dynamics, with area
PROT. VEG.	unconstrained dynamics, with area

Final Form Model

HORTICULTURE	MNL model. Redefined, excluding Orchard area unconstrained dynamics with area OLS model. For Total Orchard area only.
ORCHARD	unconstrained dynamics, without area
SOFT FRUIT	unconstrained dynamics, with area
VEGETABLES	unconstrained dynamics, with area
PROT. VEG.	unconstrained dynamics, with area

It is clear from a comparison of these two sets of results that the imposition of stability on the model has resulted in a substantial loss of within period performance, with the Orchard sector being most affected. In an effort to overcome this it was decided to remove the Orchard sector from the top level model, and use a simple ad-hoc OLS equation for it instead. The Orchard sub-model was retained, to allocate the total between the Hard and Soft Orchards.

Thus the only change to the model is that 'Horticulture' (Area 60 in Table 2 above) is now defined as the sum of Soft Fruit, Vegetables and Protected Vegetables (A42 + A51 + A47). The model structure can be represented as:

Figure 2 The "Adjusted" Full Model

```
                                         Horticulture (60)
                                ___________________|____________________
                               |                   |                    |
                               |                   |                    |
Orchard(50)           Soft Fruit(42)        Vegetables(51)        Prot.Veg(47)
     |                 /   |   |   \               |               /   |   |   \
     |               (7) (8) (9) (10)              |           (31) (32) (33) (34)
  ___|___                                 _________|_______
 |       |                               |     |     |     |
(40)   (41)                            (43)  (44)  (45)  (46)
```

The estimation results of this new specification (for the period 1965 to 1982) are reported in Appendix 1, and the simulation results are given in the third column of Table 3, under the heading of the "Final Form Model". It will be noted that the performance has been improved, not only for the Orchard sectors, but in most of the others also. It is this final specification which will be used when the returns are made endogenous, and it is to this that we now turn.

Specification of the Returns per Hectare Equations

For most crops, the modeling of the returns per hectare was done in several stages. The price equations were estimated in double log form, and generally had the following structure.

$$Ln(P_i) = r_1 + r_2.Ln(TPDI) + r_3.Ln(A_i.Y_i) \qquad 10)$$

where TPDI is Total Personal Disposable Income and Y the crops gross yield (i.e. the total available for harvest, rather than the quantity actually harvested. This avoids the complication of the price and net yield being simultaniously determined).

For some crops it was thought that the output of competing crops may affect the price, and so the relevant variables were included also.

A feature of the Horticultural sector is that in some years all of the output that is available for sale is not sold, due to poor quality or a glut of produce. It was therefore decided that an output harvested equation should be estimated, of the form

$$Ln(OH_i/(A_i.Y_i)) = o_1 + o_2Ln(Y_i) \qquad 11)$$

where OH is the output harvested, and the dependant variable is the proportion of gross output ($A_i.Y_i$) that is harvested. The most significant determinant of this is the yield level, so that in years of high yield the proportion of gross output harvested is low. For some crops the yield was not significant, and in those cases the mean of the dependent variable is used.

Using equations 10 and 11 above as an example, the log of returns per hectare can now be determined as

$$Ln(P_i.OH_i/A_i) = r_1 + r_2.Ln(TPDI) + r_3.Ln(A_i.Y_i) + o_1 + (1+o_2).Ln(Y_i) \qquad 12)$$

For some crops, this procedure was not possible. This is because some crops are aggregates of a number of diverse sub crops (for example, 'others'(46) in the vegetable sector), and so one cannot define an aggregate quantity produced. In those cases, the returns per hectare were estimated directly. It is not clear cut as to which explanatory variables should be used in such an equation but the preliminary investigation suggested that the following specification worked quite well.

$$Ln(RET_i) = r_1 + r_2.Ln(TPDI) + r_3.W + \Sigma_{j=1}^{k} r_{4j}.Y_{ij} \qquad 13)$$

where W represents the weather variables relevant for a particular sector, and Y_{ij} is the yields of a subset of the crops that make up the sub sector i. The equations estimated for each crop or aggregate group are reported in Appendix 2.

It is intended that there shold be further development of the returns sector of the model. If the model can be extended to the full 33 crop specification the problems caused by using aggregate sub sectors will be overcome. Until that is possible, it is thought that the aggregate returns (e.g. for brassicas) may be constructed as weighted average of the lower level returns, where the weights used are the average areas of the crops, rather than the actual areas which should be used (and which cannot be because the model does not disaagregate down to that level). On a more general level, it is intended to expand the price equations, so that the impact of other factors, such as imports, can be included.

It will be noted that no attempt has been made to explain the yields of the individual crops, so that in the simulations reported below they are treated as exogenous variables. The reason for this is that it is thought that the major determinant of yields is the weather, and therefore, if the model is to be used for forecasts of future developments in the horticultural sector, then average weather would have to be used, and therefore average yields generated. The only case were this is not true is if there is a trend in the yield, when it may be necessary to estimate a full yield equation in order to be able to accurately extrapolate the trend of the yield. This is only the case with protected vegetables, and yield equations for those crops will be developed if time permits.

Simulation with Endogenous Returns

Having estimated the returns equations for the lowest level crops it is then possible to simulate the full model, with returns endogenously determined. It should be noted that the higher level returns per hectare (needed in the top level sub-model) are also generated within the model, and consist of weighted averages of the relevant lower level returns, where the weights used are the areas to each of the lower level crops. The U(2) statistics are given in Table 4 below in the first two columns. As is to be expected, the results are not as good as when the returns are exogenous, but are still very acceptable.

In order to close the model it is only left to determine the total area in Horticulture, as up to now this has been taken as exogenous, and the model simply allocates this area between the different activities. Two possibilities have been considered within the context of the Manchester Model. The first is to construct a further MNL model that would allocate some higher area (for example, cultivated land) between competing activities (e.g. cereals, rape etc), one of which would be Horticulture. However, given the problems associated with the higher level model within horticulture, it was thought more prudent to take an ad-hoc approach and specify a single equation that determines the horticultural area. The estimated equation is given in Appendix 1, but the general form of the equation is to use a lagged dependent, lagged returns to horticulture deflated by an index of labour costs, and lagged returns to wheat deflated by an index of fertilizer costs. The inclusion of this equation into the system means that the exogenous variables needed to run the model are relatively few. Most of these are outside the bounds of what one may describe as the Horticultural sector, but some may be determined in other sub models within the overall Manchester Model. The full list of exogenous variables contains weather variables, yields of some Horticultural crops, Wage and Fertilizer price indices, Wheat price index and Wheat yields and Total Personal Disposable Income.

The U(2) statistics for this complete system are also given in Table 4. These are also very acceptable, and for only one crop (vegetables) do the U(2) statistics show a marked increase over those generated when the total area is exogenous. The performance of the total area equation is also good, given that the returns to total horticulture are generated within the model at a much lower level, and then aggregated up.

4) THE TRUNCATED MODEL

The model we have been dealing with so far is fairly large, with some 60 equations, and that is without the accounting equations needed to generate total revenue (see section 5). It was thought that this may be too large for inclusion in the full Manchester Model, and so a 'Truncated' model has been developed. It is envisaged that this reduced model will be used in general simulation runs, but that the full model may be used if there is a particular interest in the Horticultural sector.

The Truncated model is simply the top levels of the full model i.e. the total area equation, the allocation of that area between Soft Fruit, Vegetables and Protected Vegetables, and the equation for Orchard area. What is now needed are equations for the returns per hectare for the four aggregate commodities. A similar approach to that used to derive the 'aggregate crop' return equations in the Full model has been used. It is a fairly eclectic approach, with the emphasis on achieving a good fit rather than consistency between equations. The equations are in double log form, with TPDI capturing the general increase in nominal returns. Other explanatory equations include the yields of important crops that make up the aggregate, the aggregate's land area, weather variables and (for the protected vegetables) the level of Tomato imports. Detailed results are given in Appendix 3.

Table 4 U2 Statistics, Dynamic Simulations 1965 to 1982

FINAL FORM MODEL, ENDOGENOUS RETURNS

	A60 Exogenous		A60 Endogenous	
	AREA	RETURNS	AREA	RETURNS
TOTAL AREA				
Horticulture	--	--	0.7654	0.4022
HORTICULTURE				
Orchard	0.5477	0.6618	0.5477	0.6618
Soft Fruit	1.0297	0.6397	0.9047	0.6368
Vegetables	0.0802	0.5092	0.7848	0.5066
Prot. Veg.	0.6063	0.6986	0.7709	0.6265
Orchard				
Hard Orchard	0.6337	0.6878	0.6337	0.6878
Soft Orchard	0.3516	0.5616	0.3516	0.5616
Soft Fruit				
Strawberries	1.1742	0.7478	1.0543	0.7478
Raspberries	0.9211	0.4874	0.9563	0.4874
Blackcurrants	0.6340	0.8268	0.6236	0.8511
Others	0.6488	0.6268	0.9161	0.6381
Vegetables				
Roots	0.4291	0.5983	0.7928	0.5983
Brassicas	0.9029	0.6332	0.9091	0.5942
Legumes	0.6788	0.5898	1.0021	0.5898
Others	1.1193	0.5251	1.1614	0.5251
Protected Veg.				
Tomatoes	1.0402	0.5741	1.0543	0.5741
Cucumbers	0.9234	0.5654	0.9563	0.5654
Lettuce	0.7705	0.7460	0.6236	0.7460
Mushrooms	0.9525	0.7067	0.9161	0.7016

The resulting model is relatively small, with 18 equations. The simulation results generated by the truncated model are given in Table 5 and the relevant values for the Full model are repeated. The comparison brings up some interesting points. In the truncated model, the returns generally have the smaller U(2) statistics, implying that the aggregate returns equations are better than the aggregation of individual return equations. However, this advantage in the returns is not translated into a similar result in the area simulations, where the Full model is better for 3 out of the 5 sectors. These differences are not large, however, and it appears one loses little at the aggregate level by using the truncated model. One obviously loses the detail of what is happening <u>within</u> the aggregates.

Table 5 COMPARISON OF RESULTS FROM THE TRUNCATED MODEL AND THE FULL MODEL

U(2) Statistics. Dynamic Simulations 1965 to 1982

	TRUNCATED MODEL		FULL MODEL	
	AREA	RETURNS	AREA	RETURNS
Horticulture	0.8092	0.3420	0.7654	0.4022
Orchard	0.5466	0.5915	0.5477	0.6618
Soft Fruit	0.8834	0.5535	0.9047	0.6368
Vegetables	0.8331	0.4631	0.7848	0.5066
Prot. Veg.	0.8722	0.4378	0.7709	0.6265

5) SIMULATION OF VALUES

So far the model has been dealing with the area and returns to the various sectors. What is needed for the Manchester Model are estimates of the value of output for Horticulture. To generate these is fairly straight forward, as we have returns per hectare and the area of each crop. The product of these will give us the value for a particular crop, and thus by aggregation, for a particular sub-sector and for Horticulture as a whole. However, some accounting adjustments have to be made. Firstly, value is needed in Calendar years, while we have to date been operating with Harvest years, which run from approximately June to May. This is not a great problem as the harvest period for many crops lies within a single calender year, i.e. the 1978/9 harvest year for Runner Beans falls completely within 1978. However for some, (notably in the vegetable sector) the calender year contains sales from two harvest years. This was dealt with in the following way. For the four groups Orchard (50) Soft Fruit (42) Vegetables (51) and Protected Vegetables (47) the value of output in calender years was calculated from Basic Horticultural Statistics. This was then regressed against the value of output for the two harvest years that fall within that calender year. This procedure effectively allocates the revenue generated in a harvest year between the two calender years that it falls in. In fact, it was only for Vegetables that any significant effect was found, with a suprisingly high proportion of the value of the Harvest year falling in the new year. There was no effect for the Orchard sector, which is surprising given the seasonal pattern of output, but that effect could not be found in the revenue figures.

Secondly, some elements of the sector have been excluded from the analysis, notably the non-edibles, but also some minor crops within both fruit and vegetables. These were incorporated on a simple % basis. Thus, the values generated by aggregating the calendar values for the crops identifed in table 2 were compared with the reported values in the Output, Input and Net Farm Income table of the Annual Review. This was done for two sub groups, All Vegetables, (47 and 51 in table 2 above) and All Fruit (50 and 47). There was no time trend evident in the relationship, and so a simple % mark up was used, of 17% for All Vegetables, and 9% for All Fruit. This simply means that the value of All Vegetables reported in the Annual Review is on average some 17% higher than the value of the vegetables (both protected and field) included in Table 2.

A similar method was used to incorporate the non-edibles into the model. The value of non-edibles was expressed as a % of the value of Vegetables Plus Fruit (as reported in the Annual Review). This had an average of value of approximately 23%, but also showed a significant upward trend over the period, which was included.

With these accounting equations included, it is now possible to simulate the model, and generate an estimate of the 'Horticultural Value', as defined in the Annual Review. This has been done for the period 1976 to 1982, for both the Full model and the Truncated model, and the results are reported in Table 6 below. Percentage errors are reported in brackets. It is interesting to note that on the basis of the Root Mean Squared Error the truncated model is better. This may reflect the fact that the returns are more important in determining value, rather than area. However, using either model, the size of the errors are acceptably small, especially for a dynamic simulation over a 7 year period.

Conclusions

This paper has reported the development of an econometric model of the U.K. Horticulture sector, a sector that has not previously been analysed in this way. The model has encompassed the area planted to particular crops as well as the prices received for the products, and the output harvested. The primary purpose of the model has been to generate the Value of Horticultural Output, for use in the model of U.K. agriculture currently under development at Manchester. When used for this purpose it is likely that the "Truncated" form of the model would be implemented, but if a wider analysis of changes in the sector is needed then the full model, with its greater disaggregation, could be used. In particular, if the price equations are extended to include the influence of imports, then the model would provide a usefull vehicle for exploring the implications of Spanish and Portugese entry into the EEC on U.K. Horticulture.

Table 6 Simulations of Values, Dynamic Simulation 1976 to 1982

VEGETABLE VALUES £m CALENDAR YEARS

	ACTUAL	TRUNCATED	FULL
1976	405.1	418.2 (3.2)	425.1 (4.9)
1977	486.4	456.3 (-6.2)	448.7 (-7.7)
1978	460.4	459.6 (0.2)	447.7 (-2.7)
1979	536.4	524.4 (-2.2)	510.3 (-4.9)
1980	560.3	576.0 (2.8)	575.6 (2.7)
1981	583.5	585.7 (0.4)	595.9 (2.1)
1982	595.8	629.6 (4.8)	637.8 (6.9)

FRUIT VALUES £m CALENDAR YEARS

	ACTUAL	TRUNCATED	FULL
1976	115.9	118.1 (1.9)	121.5 (4.9)
1977	144.6	128.9 (-10.)	118.9 (-18.)
1978	152.5	163.6 (7.3)	153.7 (0.1)
1979	157.6	153.6 (-2.5)	166.4 (4.8)
1980	169.7	182.5 (7.5)	179.8 (5.3)
1981	187.2	188.5 (0.7)	199.3 (5.8)
1982	212.1	220.6 (-4.0)	208.8 (-2.5)

HORTICULTURE VALUES £m CALENDAR YEARS

	ACTUAL	TRUNCATED	FULL
1976	629.4	647.4 (2.8)	659.8 (4.9)
1977	755.1	710.3 (-5.9)	689.1 (-9.3)
1978	749.7	760.8 (1.5)	734.1 (-2.3)
1979	854.2	832.2 (-2.6)	830.5 (-2.9)
1980	912.9	936.1 (2.5)	932.3 (1.9)
1981	962.5	960.7 (0.2)	986.8 (2.3)
1982	1012.4	1054.6 (4.9)	1056.4 (4.0)
		RMS ERROR 27.4	RMS ERROR 35.3

APPENDIX 1

Parameter Estimates

Parameter estimates generated by the MNL model are difficult to interpret, as they are in mean deviation form, and therefore a parameter that is insignificant from zero does not imply that the variable should be excluded, but that the variable has an equal effect across all equations. This Appendix reports the results for each of the five sub models. In order to simplify the presentation, some conventions of notation should be noted. Individual crops are identified by their number in Table 2. LWWn refers to the log of the normalised share for crop n. The presence of .1 implies a one year lag. LNRETnx denotes the log of the ratio of returns to crops n and x. LNAn is the log of the area n. 't' statistics are not reported for the MNL model as they give little information about the importance of a variable in a particular regression. All equations have been estimated over the period 1964 to 1982, using annual data.

Area Model Parameter estimates

Total Horticulture Area
(t stats. in parentheses)

$$A60 = \underset{(2.41)}{80470} + \underset{(4.11)}{0.553}\ A60.1 + \underset{(2.73)}{639624}\ RET60.1/WAGE.1 - \underset{(2.91)}{6028}\ WHEATRET.2/FERTP.2$$

R BAR SQRD.	=	0.724
F TEST	=	15.8
D.h.	=	-1.24
d.f.	=	14

Horticulture sub model

Dependent Variable

	LWW51	LWW42	LWW47
Intercept	-4.60	7.018	-2.41
LWW42.1	0.193	0.581	-0.773
LWW51.1	0.744	0.149	-0.893
LNRET5142.1	0.132	-0.145	0.0134
LNRET4742.1	-0.0427	0.0754	-0.0327
LNA60	0.441	-0.632	-0.191

Total Orchard area

(t stats. in parentheses)

$$A50 = \underset{(2.61)}{42514} + \underset{(1.89)}{0.404}\ A50.1 + \underset{(1.61)}{39322}\ RET50.1/WAGE.1 - \underset{(2.72)}{903}\ TIME$$

R BAR SQRD.	= 0.988
F TEST	= 480
D.h.	= 3.53
d.f.	= 14

Orchard sub model

	Dependent Variable	
	LWW40	LWW41
Intercept	2.11	-2.11
LWW40.1	0.647	-0.647
LNRET4041.1	0.0112	-0.0112
LNA50	-0.166	0.166

Soft Fruit sub model

	Dependent Variable			
	LWW7	LWW8	LWW9	LWW10
Intercept	0.178	-2.08	-0.257	2.16
LWW7.1	0.288	0.726	-0.289	-0.725
LWW8.1	-0.00347	0.490	0.153	-0.639
LWW9.1	-0.269	-0.109	0.562	-0.184
LNRET87.1	0.0153	0.0107	-0.0584	0.0324
LNRET97.1	0.0278	-0.0583	-0.0309	0.0614
LNRET107.1	-0.0514	-0.0662	0.0946	0.0229
LNA42	0.0258	0.161	0.0466	-0.233

Vegetable sub model

	Dependent Variable			
	LWW43	LWW44	LWW45	LWW46
Intercept	-3.60	0.961	-1.35	3.99
LWW43.1	0.735	0.022	-0.127	-0.631
LWW44.1	0.264	0.948	-0.555	-0.657
LWW45.1	0.303	0.247	0.441	-0.991
LNRET4344.1	0.335	-0.168	-0.0944	-0.0725
LNRET4544.1	-0.060	0.0671	0.0329	-0.040
LNRET4644.1	-0.0685	-0.057	-0.00491	0.130
LNA51	0.269	-0.0823	0.168	-0.355

Protected Vegetable sub model

	Dependent Variable			
	LWW31	LWW32	LWW33	LWW34
Intercept	1.69	6.59	-5.54	-2.74
LWW31.1	1.25	0.193	-0.823	-0.624
LWW32.1	0.229	0.587	-0.214	-0.602
LWW33.1	0.331	0.786	-0.0758	-1.04
LNRET3132.1	0.242	0.040	-0.174	-0.108
LNRET3332.1	-0.0967	0.085	0.146	-0.134
LNRET3432.1	-0.207	-0.0721	0.00264	0.276
LNA47	-0.234	-0.948	0.85653	0.326

APPENDIX 2

This Appendix reports the estimated equations for the returns per Hectare, either directly, or through separate price and output harvested equations. A list of variable names is in Appendix 4, but one general point will be made here. The Output Harvested equations are estimated with the dependent variable defined as the log of the ratio of output harvested to gross output (e.g. LNOH%7). At times this ratio is very constant over time, which is why the apparent fit is so poor. In fact for most crops the determination of Output Harvested is quite high.

Returns Equations

Strawberries

$$\text{LNP7} = \underset{(15.4)}{2.22} + \underset{(25.9)}{0.894}\ \ln(\text{TPDIH})$$

R BAR SQRD = 0.974
F TEST = 674
D.W. = 1.44
d.f. = 17

LNOH%7 = -6.97

mean value used

Raspberries

$$\text{LNP8} = \underset{(9.59)}{1.81} + \underset{(20.8)}{0.941}\ \ln(\text{TPDIH})$$

R BAR SQRD = 0.959
F TEST = 431
D.W. = 1.37
d.f. = 17

LNOH%8 = -6.95

mean value used

Blackcurrants

$$\text{LNP9} = \underset{(15.4)}{2.22} + \underset{(25.9)}{0.894}\ \ln(\text{TPDIH})$$

R BAR SQRD	= 0.974
F TEST	= 674
D.W.	= 1.44
d.f.	= 17

$$\text{LNOH\%9} = \underset{(102)}{-6.82} - \underset{(2.2)}{0.00173}\ \text{SERAIN\$JUL}$$

R BAR SQRD.	= 0.18
F TEST	= 4.94
D.W.	= 2.14
d.f.	= 17

Others

$$\text{LNP10} = \underset{(1.86)}{5.5} + \underset{(12.1)}{0.874}\ \ln(\text{TPDIH}) - \underset{(1.39)}{0.4}\ \ln(\text{A10.Y10})$$

R BAR SQRD	= 0.952
F TEST	= 178
D.W.	= 1.12
d.f.	= 16

$$\text{LNOH\%10} = \underset{(107)}{-6.75} - \underset{(3.1)}{0.115}*\ln(\text{Y10}) + \underset{(3.72)}{0.00068}\ \text{SERAIN\$JUL}$$

R BAR SQRD.	= 0.475
F TEST	= 9.16
D.W.	= 1.97
d.f.	= 16

Tomatoes

$$\text{LNP31} = \underset{(3.6)}{14.4} + \underset{(5.08)}{1.051}\ \ln(\text{TPDIH}) - \underset{(2.5)}{1.92}\ \ln(\text{Y31}) - \underset{(2.5)}{0.764}\ \ln(\text{IM31})$$

R BAR SQRD	= 0.958
F TEST	= 139
D.W.	= 1.21
d.f.	= 15

$$\text{LNOH\%7} = \underset{(119)}{-6.65} - \underset{(4.9)}{0.058}*\ln(\text{Y31})$$

R BAR SQRD.	= 0.57
F TEST	= 24.9
D.W.	= 2.76
d.f.	= 17

Cucumbers

LNP32 = 3.15 + 0.498 ln(TPDIH)
(20.6) (13.6)

R BAR SQRD = 0.911
F TEST = 186
D.W. = 2.30
d.f. = 17

LNOH%32 = -6.03 - 0.167*ln(Y32)
(18.2) (2.7)

R BAR SQRD. = 0.261
F TEST = 7.35
D.W. = 1.32
d.f. = 17

Lettuce

LNP33 = 6.59 + 0.804 ln(TPDIH) - 1.34 ln(Y28)
(8.8) (14.9) (4.6)

R BAR SQRD = 0.937
F TEST = 135
D.W. = 1.82
d.f. = 16

LNOH%33 = -4.69 - 0.697 SERAIN$JUL
(9.2) (4.3)

R BAR SQRD. = 0.49
F TEST = 18.7
D.W. = 1.08
d.f. = 17

Mushrooms

LNP34 = 7.68 + 0.819 ln(TPDIH) - 0.455 ln(A34.Y34)
(6.34) (15.2) (3.51)

R BAR SQRD = 0.967
F TEST = 266
D.W. = 1.20
d.f. = 16

LNOH%34 = -6.70 - 0.0414 ln(Y34)
(60) (1.79)

R BAR SQRD. = 0.110
F TEST = 3.22
D.W. = 2.31
d.f. = 17

Hard Orchard

```
LNRET40  = -2.46 + 0.798 ln(TPDIH) - 0.179 ln(MRAIN$AUG)
           (3.69)  (10.9)            (1.52)

                + 0.01048 MMINT$MAY
                  (0.23)

R BAR SQRD.  = 0.927
F TEST       = 77.5
D.W.         = 1.46
d.f.         = 15
```

Soft Orchard

```
LNRET41  = -3.66 + 0.796 ln(TPDIH) - 0.269 ln(MRAIN$JUN/MSUN$JUN)
           (10.7)  (9.58)            (2.78)

                - 0.348 LN(MRAIN$AUG/MSUN$AUG)
                  (2.37)

R BAR SQRD.  = 0.863
F TEST       = 38.9
D.W.         = 1.23
d.f.         = 15
```

Roots

```
LNRET43  = -3.45 + 0.880 ln(TPDIH) - 0.141 ln(EARAIN$JUN/EASUN$JUN)
           (19.6)  (20.9)            (3.30)

R BAR SQRD.  = 0.962
F TEST       = 231
D.W.         = 1.49
d.f.         = 16
```

Brassicas

```
LNRET44  = 11.5  + 0.762 ln(TPDIH) - 0.103 ln(EARAIN$JUN/EASUN$JUN)
           (2.14)  (13.9)            (3.01)

                - 0.01048 ln(A44)
                  (2.79)

R BAR SQRD.  = 0.976
F TEST       = 243
D.W.         = 2.04
d.f.         = 15
```

Legumes

LNRET45 $-3.43 + 0.638 \ln(TPDIH)$
(26.6) (20.7)

R BAR SQRD. = 0.959
F TEST = 4285
D.W. = 2.38
d.f. = 17

Others

LNRET46 = $-2.69 + 0.862 \ln(TPDIH) - 0.123 \ln(EARAIN\$JUN/EASUN\$JUN)$
(15.3) (20.6) (2.90)

R BAR SQRD. = 0.961
F TEST = 222
D.W. = 1.97
d.f. = 16

APPENDIX 3

Estimates of the aggregate returns per Hectare equations used in the "Truncated" model

Orchard

```
LNRET50  = -2.60 + 0.802 ln(TPDIH) - 0.337 ln(MRAIN$AUG)
           (5.39)  (17.6)            (3.07)

                 + 0.165 ln(MRAIN$JUN)
                   (2.41)

R BAR SQRD.    = 0.949
F TEST         = 113
D.W.           = 1.42
d.f.           = 15
```

Soft Fruit

```
LNRET42  = 7.01  + 0.768 ln(TPDIH) - 0.727 ln(Y7) - 0.597 ln(Y8)
           (1.57)  (18.5)            (3.73)         (2.78)

                - 1.22 ln(A42)
                  (2.75)

R BAR SQRD.    = 0.982
F TEST         = 240
D.W.           = 2.30
d.f.           = 14
```

Vegetables

```
LNRET51  = -3.34 + 0.798 ln(TPDIH) - 0.084
ln(EARAIN$JUN/EASUN$JUN)
           (30.2)  (30.3)            (3.15)

R BAR SQRD.    = 0.981
F TEST         = 474
D.W.           = 1.73
d.f.           = 16
```

Protected Vegetables

```
LNRET47  = 4.30  + 0.693ln(TPDIH) - 0.196 ln(MSUN$AUG) -
0.597ln(IM31)
           (4.10)  (37.3)           (2.54)                 (3.29)

R BAR SQRD.    = 0.988
F TEST         = 493
D.W.           = 1.55
d.f.           = 15
```

APPENDIX 4

Variable Definitions

Individual Crops, or Aggregates of Crops are identified by the number given in Table 2.

Many of the variable names follow a particular classification scheme. Thus

An	Denotes the area of crop n.
LNAn	Denotes the Log of area n.
LNWWn	Denotes the normalised share of crop n within its immediate grouping.
Yn	Denotes the Yield per Hectare of crop n.
RETn	Denotes the Returns per Hectare to crop n.
LNRETnx	Denotes the log of the ratio of Returns per Hectare to crops n and x.
LNPn	Denotes the log of Price per tonne of crop n.
OHn	Denotes the Harvested Output of crop n.
LNOH%n	Denotes the log of the ratio of Harvested Output to Gross Output for crop n
TPDIH	Total Personal Disposable Income, in Harvest Years.
SERAIN\$JUL	Rainfall in the South East in July, as a % of Monthly Average.
MRAIN\$AUG	Rainfall in the Midlands in August, as a % of Monthly Average
MRAIN\$JUN	Rainfall in the Midlands in June, as a % of Monthly Average
MSUN\$JUN	Hours of Sunlight in the Midlands in June, as a % of Monthly Average
MSUN\$AUG	Hours of Sunlight in the Midlands in August, as a % of Monthly Average
EASUN\$JUN	Hours of Sunlight in East Anglia in June, as a % of Monthly Average
EARAIN\$JUN	Rainfall in East Anglia in June, as a % of Monthly Average
MMINT\$MAY	Minimum Air Temperature in the Midlands in May, Degrees Centigrade, constrained to equal zero if positive.
IM31	Imports of Tomatoes.

WHEATRET Returns per Hectare to Wheat.

FERTP Fertilizer Price Index.

WAGE Wage Index.

TIME Time Trend.

References

Bewley,R., D.R.Colman and T.Young — A System Approach to Modelling Supply Equations in Agriculture. JAE Forthcoming

Bewley,R. and T.Young — Applying Theils Multinomial Extension of the Linear Logit Model to Meat Expenditure Data Amer. Jour. of Agric Econ. Forthcoming.

MAFF Statistics, (Agricultural Commodities)B — Basic Horticultural Statistics for the United Kingdom. Various Years

Horler,E — The Mini Produce Manual.