Modeling adoption of innovations in agriculture using discrete choice models

Daniel Shefer, Mordechai Cohen, Shlomo Bekhor
Technion- Israel Institute of Technology, Haifa, Israel

Abstract

This paper is concerned with the development of varieties and fertilization techniques of greenhouse tomatoes, and their spatial diffusion in the northwestern region of the Negev in Israel. The main objective of the paper is to identify the factors affecting the farmers' decision to adopt innovations and the factors inducing the process of knowledge-diffusion in the rural region. The approach adopted is the use of discrete choice models based on random utility theory.

Results of the empirical analysis when applying the disaggregate Logit Model indicate that the regional, local and individual attributes have a significant bearing on the farmers' decision-making process in regard to choosing among alternative tomato varieties and fertilization techniques. The findings indicate that the models constructed for this study may be used as a planning tool for the purpose of evaluating the effect of different factors on the spatial diffusion of innovations in rural regions. The results of the research could also assist decision-makers in formulating development policies for rural regions.

Keywords: Spatial diffusion; discrete choice models; greenhouse tomatoes; nested logit

Introduction

This paper is concerned with the development of an agricultural product and the spatial diffusion of innovations as an important element in the economical development of a rural region. Specifically, the paper considers the development greenhouse tomatoes, and their spatial diffusion in the northwestern region of the Negev in Israel. Two aspects of difusion of innovation will be dealt with in this paper. First, the selection of a specific tomato variety by the farmers, and the second is the choice among alternative fertilization techniques.

Many studies have dealt with questions related to the spatial diffusion of innovations in the manufacturing industries and their relation to regional development. Discrete choice models have been used in many studies related to the manufacturing sector: Frenkel et al. (2003) is a recent example. However, in the agricultural sector the use of discrete choice models in the innovation context is rare. This research aims to mimic the process observed in the manufacturing sector, regarding sophisticated and advanced farming practices.

The spatial diffusion of innovations has been studied by many researchers who were concerned with urban and regional development. Most of the recent studies focused on the industrial sector, although in the past other studies were concerned with the agricultural sector. Examples of past studies can be found in Mansfield (1968), Yeats (1974), and Rogers (1983). These researchers studied the process of generating innovations and their spatial diffusion, and the consequent economic and social impacts.

The mathematical models used to describe diffusion processes assume that the new products do not change throughout the adoption period. However, many studies showed on-going product improvements together with new product generation, competition and changes in the production processes in the adoption period. Examples of these studies can be found in Metcalf (1981), Kamien and Schwartz (1982), and Davelaar (1991). Nevertheless, there seem to be a consensus among researchers that product lifecycle can be distinctly described by several phases over time: diffusion, growing, mature, saturation and decay. The last phase coincides with the appearance of a new alternative or impruved product, which is preferred and superior to the first one. Examples of such processes can be found in Rogers (1983) and Staudt and Taylor (1965).

In addition, geographical aspects have long been recognized as important factors in the spatial diffusion of innovations, as can be seen in the studies of Hagestrand (1967), and Meir (1981).

In the agricultural sector environment plays an important role, as one should expect. In most countries, agricultural production is split among many small producing units; these units are generally spread over a large area, and therefore are located far from urban centers and services associated with them. For this reason, the contribution of R&D centers situated close to the agricultural farms may be quite significant and detrimental, as found in several studies: see, for example, Griliches (1957), Evenson and Kislev (1975), and Arndt et al (1977).

The main objective of the paper is to identify the factors affecting the farmers' decision to adopt innovations and the factors inducing the process of knowledge-diffusion in the rural region. The approach adopted is the use of discrete choice models based on random utility theory.

This study purposely concentrates on a specific area (in this case, the northwestern region of the Negev in Israel). It is a relatively new region, where the government encouraged farmers to grow export-oriented products. The greenhouse tomato was selected to be the proper crop for that purpose. Concomitedly, it was decided to establish in the region an agricultural R&D center, with the purpose to augment the farmers' ability to cope with the tough competition of exporting tomatoes, particularly to Europe but also to North America.

The paper is divided into four sections. The next section deals with the theoretical background and the research methodology. Empirical results of disaggregate logit models are presented in the subsequent section. Summary and conclusions form the last part of the paper.

Theoretical Background

The models derived in this paper are based on the perceived attractiveness of each of the available alternatives. In the present study, the alternatives are defined as the different tomato varieties. The available alternatives are expressed as weighted sums of attributes; in our case, the attributes can be classified as regional, local or personal attributes. The measure of attractiveness of an alternative is referred to as its 'utility'. The basic theory used in the derivation of the models is the discrete choice - random utility theory. In this paper we presents only a brief summary of this theory. For a more comprehensive review, see Ben-Akiva and Lerman (1985).

We assume that each farmer perceives the utility associated with each brand of tomato variety available, and chooses the one with the greatest perceived utility. We assume that the utility U_{in} of alternative *i* for individual *n* can be decomposed in two terms: a deterministic term V_{in} , which is associated with the measured attributes of the alternative, and a random error e_{in} representing the difference between the measurable utility and the true utility of the alternative for individual *n*. The random error accounts for factors affecting the utility of an alternative not included in V_{in}, as well as other factors, which are fundamentally unobservable.

The probability that individual *n* chooses alternative *i* from the set of available alternatives J_n is equal to:

$$P_n(i) = P(U_{in} \geq U_{jn}, \forall j \in J_n) = P(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j) \quad (1)$$

Rearranging the terms:

$$P_n(i) = P(\varepsilon_{jn} - \varepsilon_{in} \geq V_{in} - V_{jn}, \forall j \in J_n) \quad (2)$$

Thus, the probability that a particular alternative is chosen depends on the joint distribution of the differences between the error terms. Several different models were developed, according to the distribution assumed for the error terms. The most common models are the probit model and the logit model. The probit model is obtained by assuming that the random terms in the equation above are normally distributed. The logit model is obtained by assuming that the random terms are independent and identically distributed according to the negative double exponential (also known as Weibull or Gumbel) distribution. The functional form of the multinomial logit model is as follows:

$$P_n(i) = \frac{\exp(V_{in})}{\sum_{j \in J_n} \exp(V_{jn})} \quad (3)$$

Probit and logit models are well known. The probit model does not have a closed functional form, and for this reason the estimation procedure becomes more complicated. The logit model has a simple functional form, and the parameter estimation is relatively straightforward. Another interesting point stems from the distribution properties of both models. The sum of normally distributed variables is also normally distributed, but the maximum of normally distributed variables can only be estimated by approximation. On the other hand, the maximum of Gumbel distributed variables is also Gumbel distributed. It is possible to interpret this property as follows: since the error terms account for unobservable attributes, the modeler (that estimates the parameters with a set of observable attributes) tries to maximize the difference between the random components, instead of simply summing them up.

The multinomial logit model is suitable for alternatives that can be unambiguously defined. When the alternatives have similarities, or when the decision process is made in a conditional way (which is often the case), the simple logit model is not suitable. For example, if we just want to model the choice to grow a specific tomato variety, the multinomial logit may suffice. However, if we want to model the choice to grow a tomato variety conditioned to a specific fertilization technique, the multinomial logit model cannot take into account this hierarchy of the decision making process. The nested logit model, which is more general than the multinomial logit model, can take into account the above feature.

The nested logit model was developed by assuming that the error terms are not independently distributed: i.e., the alternatives are correlated. The model is based on the assumption that the alternatives from the choice set can be divided into mutually exclusive and collectively exhaustive groups (nests) m in such a way that the error term is represented as the sum of the group-related e_m and alternative-specific e_i components, where the group-related component expresses the similarity between the alternatives. The analytical form of the nested logit model is presented as follows:

$$P(i) = \frac{\exp\left(V_i + (\mu - 1)\ln \sum_{k \in J_m} \exp(V_k)\right)}{\sum_{j \in J} \exp\left(V_j + (\mu - 1)\ln \sum_{k \in J_m} \exp(V_k)\right)} \qquad (4)$$

Where μ is a scale parameter, which represents the nesting coefficient. To be consistent with random utility theory, this parameter should lie between 0 and 1. When μ is equal to 1, the model collapses to the multinomial logit model.

The next section of the paper presents the explanatory variables used in the different models, and estimation results of the parameters in the different models.

Methodology and Data Sources

The research was intentionally performed in a region where diffusion of new technologies could be traced and observed. The Northwest region of the Negev (Southern Israel) was selected for this study, because of the high concentration of agricultural settlements in that region. In addition, a regional Research and Development (R&D) agricultural center was established there by the government. This center provides services and incentives for the farmers to grow new varieties and emloy advanced fertilization techniques. The greenhouse tomato research activities were initially developed in this area, in which new varieties were developed and tested. The successful new technologies tested were later mimicked by other rural regions in Israel.

This research gathered data from two main sources. The first source was aggregate data on the agricultural settlements, such as socio-economic indicators. The second source was a survey performed among 151 farmers from 21 different agricultural settlements in the region. This survey collected data at the individual level, which formed the basis for the empirical models tested in this study. A detailed description of the methodology and data sources can be found in Cohen (1997).

The first step of the study concentrated on the historical development of agriculture techniques in the region. The information on different tomato varieties was collected, as well as the different fertilization techniques used in greenhouses. The next step was to perform statistical analysis on the data collected. The analysis at this stage focused on macro-economic indicators. The main step of the analysis, which is described in detail in the next sections, was the analysis perform at a disaggregate level, i.e. at the individual level.

The main hypotheses tested in this paper are described as follows:

- The interaction between the agricultural settlements and the R&D regional center not only contributes to the initiation of innovations, but also accelerates the innovation diffusion process. The variable that will represent this interaction is the distance of the settlement from the R&D center.

- The innovations and improvements developed in the R&D center will diffuse faster if they are accompanied by incentives such as financial help, guidance, etc.
- In a system composed of agricultural production units supported by a dynamic R&D center, the adoption of an innovation is not a single decision, but rather a choice to test new varieties with respect to existing ones.
- The choice of an alternative is dependent on attributes of the farmer, such as age, education, place of origin, etc.

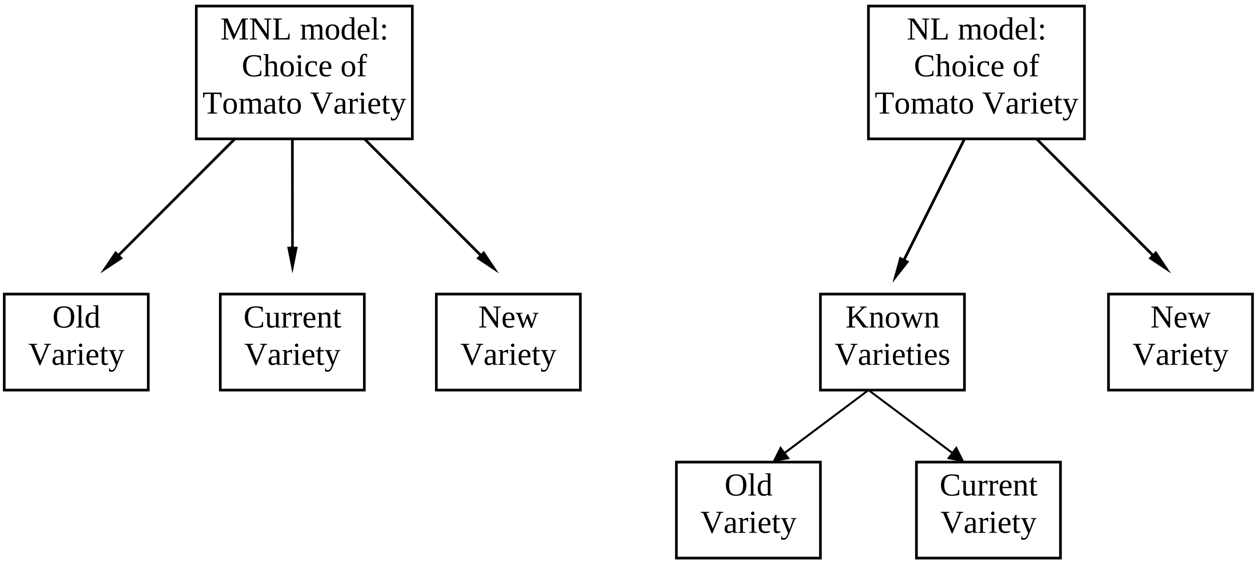
Model Estimation

Discrete choice models will be used to estimate the parameters that influence the choice between alternative tomato varieties and fertilization techniques. Specifically, the Multinomial Logit (MNL) and the Nested Logit (NL) models will be used as basis for the estimations.

The MNL and NL models are commonly used in disciplines such as transportation (choice among transport modes, for example), or market research (choice among beverages, for example). In such disciplines, practitioners have a good deal of information on the main variables that influence the model. This is not the case in the current study. For this reason, much effort was placed on the investigation of the relevant variables that may affect the choice of new technologies. In order not to disqualify any variable that might be significant, several trials were performed, where each of them combined many of the possible variables collected in the survey. The paper presents the results of the selected models used in the analysis, which captured the main variables that significantly influenced the choice among the alternative technologies.

It is possible to represent the multinomial logit and the nested logit in a tree diagram, as depicted in Figure 1.

Figure 1. Tree Diagram of MNL and NL models



The figure above illustrates a hypothetical choice between 3 tomato varieties (old, current and new), which could be modeled in two different ways. The MNL model assumes that each alternative is distinct from one another, and then each alternative is in the same level. In the NL model, the old and current varieties are assumed to have a certain amount of correlation (named "known" varieties), and then are grouped in a common nest.

In the present study, there are two possible alternative sets: (1) the different tomato varieties and (2) the different fertilization techniques. We present the estimation results for 3 different models: the first model is related to the choice of tomato varieties, the second is related to the choice of fertilization techniques, and the third is a mix of fertilization techniques and tomato varieties. The following sections describe the results of the estimations for each model. Note that for each model, several sets of variables were tested, but we elected to show only the models with the best fit.

Results

Model 1

The first model tested was the MNL model that estimates the choice among different tomato varieties. The varieties were grouped into three categories related to the tomato varieties cultivated in 1995. The first corresponds to old varieties (121 and before), the second correspond to the current, most common variety (144), and the third corresponds to the new varieties (from 175). Table 1 below summarizes the estimation results.

Table 1: Model 1 – Choice between tomato varieties - Estimation Results

Variable	Alternative	Coefficient	T-value
Constant	1	1.511	1.7 *
Constant	2	4.289	5.5 **
Percent Export	1, 2	0.083	1.8 *
Greenhouse area ratio: farmer / total settlement	2, 3	11.37	1.6 *
Dummy variable: propensity to innovation	3	2.757	3.3 **
Dummy variable: use of new fertilization technique	1	-0.709	-1.6 *
Settlement age (years)	1	0.070	3.9 **
Distance to R&D center (km)	2	-0.069	-4.4 **

* Significant at 0.1 level

** Significant at 0.05 level

Total number of observations: 148

Log-likelihood (null model): -162.59

Log-likelihood (constants only): -128.51

Log-likelihood (final model): -95.88

Rho-bar squared w.r.t. null model: 0.41

Rho-bar squared w.r.t. constants: 0.25

The second column of the table indicates the relationship between the coefficient estimates to the utility of each alternative. Since the probability calculation is based on the difference between alternatives, the constant of an alternative is set to zero (in this case, alternative 3). The same applies to dummy variables. In general, it is common practice to relate a dummy variable to a specific alternative, to allow for easy interpretation. The T-value of 1.6 indicates that it is possible to reject the null hypothesis that the coefficient is equal to zero at a 90% confidence level.

The following conclusions can be drawn from the estimation results:

- The positive coefficient of export percentage means that export-oriented farmers prefer to grow old and current varieties, which have a proved results.
- The ratio of farmer's greenhouse area to total settlement area coefficient indicates that farmers with high ratio will be more prone to try new varieties.
- The dummy variables exhibit expected signs. The first variable, related to the use of new technologies is negative, meaning that the utility of alternative 1 (old varieties) decreases when the farmer use new technologies. The second variable, related to the propensity of innovation, is positive, meaning that the utility of alternative 3 (new varieties) increases.
- The settlement age influences the choice of tomato variety: the older the settlement, the higher is the utility to use old varieties.
- The sign of the distance to R&D center may be explained as follows: the current (most common) variety was developed and stimulated by the R&D center some years ago, and therefore settlements close to the R&D center were exposed first to this variety.

Model 2

The second model tested is related to the choice between fertilization techniques. Note that in the first model, the technique was a dummy variable. This model represents an alternative way of modeling innovation, in which the fertilization technique is the choice, and the tomato variety is an explanatory variable. This dichotomy (independent versus dependent variable) may occur also in other disciplines. For example, to model choice between transport modes, a possible explanatory variable is auto ownership. Alternatively, to model auto ownership, the transport mode may serve as an explanatory variable.

This model has two alternatives. The first alternative comprises existing fertilization techniques such as hormones, mechanical bees or ventilation. The second alternative is a new fertilization technique based on Combo Bees inside the greenhouse. Table 2 presents the estimation results.

Table 2: Model 2 – Choice between fertilization techniques - Estimation Results

Variable	Alternative	Coefficient	T-value
Constant	1	4.464	2.6 **
Settlement Age	2	0.037	2.7 **
Greenhouse area ratio: farmer / total settlement	2	9.918	2.0 *

Dummy variable: consult R&D center	2	0.759	1.8 *
Dummy variable: grow new tomato varieties	2	0.029	1.9 *

* Significant at 0.1 level

** Significant at 0.05 level

Total observations: 150

Log-likelihood (null model): -103.07

Log-likelihood (constants only): -103.97

Log-likelihood (final model): -94.04

Rho-bar squared w.r.t. null model: 0.10

Rho-bar squared w.r.t. constants: 0.10

The following conclusions can be drawn from the estimation results:

- Contrary to the first model, the settlement age influences positively the use of new fertilization techniques.
- The dummy variables exhibit expected signs.
- In both models, the greenhouse area coefficient has the same order of magnitude and sign. This means that the relative strength of the farmer in the settlement positively influences the use of new fertilization techniques.
- The constant value is relatively high compared to other variables.
- Overall measures of fit (final likelihood and rho-bar squared) for this model are inferior to the measures obtained in Model 1.

The overall conclusion from these two models is that farmers are more sensitive to choice between tomato varieties than choice between fertilization techniques. Nevertheless, variables that account for the adoption of innovation are significant in both models.

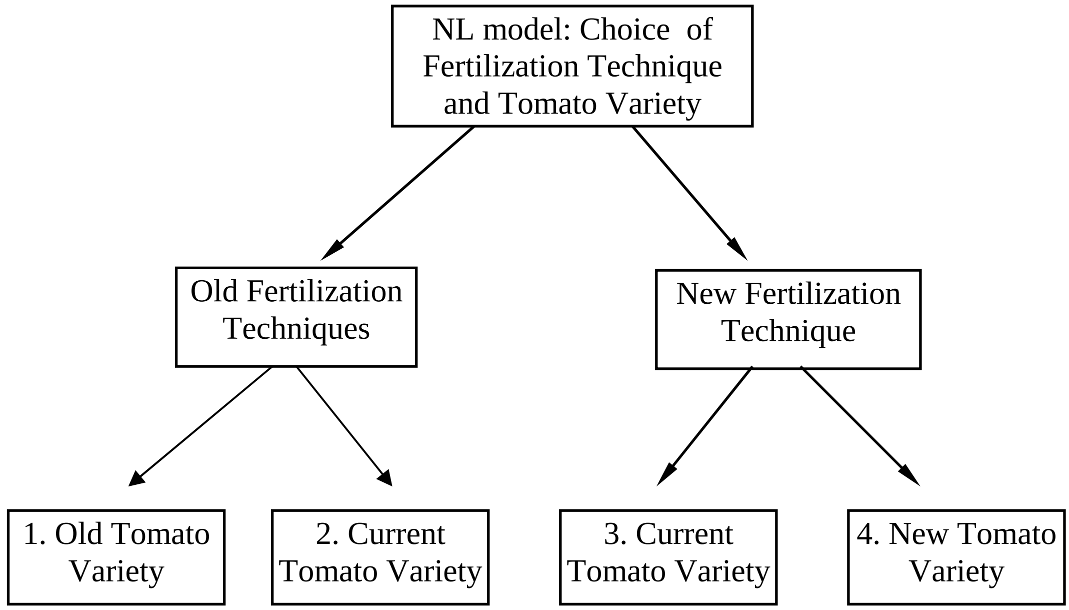
Model 3

This model combines the features from the two models presented above. The first and second models were modeled using simple MNL models. The third model indicates that the choice of fertilization technique is conditioned on the choice of the tomato variety. Since we are interested in modeling the adoption of innovations in agriculture, the hypothesis is that the choice of a new fertilization technique follows the choice of the new (innovative) tomato variety.

In order to model the choice of fertilization techniques and tomato varieties, the Nested Logit model is used. The structure of the NL model may vary, and generally the modeler judgment plays a significant role. Again using a transportation example, the choice between car, bus and train may be modeled as a nesting structure between car and transit, and the transit nest includes the choice between bus and train. However, if the train alternative represents a high-speed train, the model could be differently represented: the nesting structure could be "attractive" modes, which would include the car and train in the same nest.

It should be noted that the nesting structure could be formed in different ways, each of them resulting in a different model. In the case of plant techniques and tomato varieties, several nesting structures were tested. The model that gave the best fit was estimated with the following nesting structure, presented in Figure 2.

Figure 2. Tree Diagram of Model 3



Note that in the Tree Diagram above the "current" tomato variety alternative appears as two independent alternatives. This is because of the similarity among farmers that cultivate the most common variety with respect to fertilization techniques. The decision process in this case is as follows: the top nest indicates the choice among fertilization techniques, and the choice of tomato varieties is conditioned on the choice of the technique. Note that at most we can construct 6 alternatives in this way; however, the correlation between old fertilization techniques and new tomato varieties is very low, as is the correlation between new fertilization techniques and old tomato varieties. For this reason, the final model contains 4 alternatives.

Table 3: Model 3 – Choice between fertilization techniques and tomato varieties - Estimation Results

Variable	Alternative	Coefficient	T-value
Percent Export	1	0.016	2.2 **
Settlement Age	1	0.042	3.8 **
Greenhouse area ratio: farmer / total settlement	2,3,4	18.80	2.9 **
Distance to R&D center (km)	3,4	-0.048	-1.6 *
Dummy variable: awareness of R&D center	3	1.558	2.4 **
Dummy variable: visit R&D center	3,4	1.413	1.8 *
Dummy variable: propensity to innovation	4	2.095	3.0 **
Ratio investment in Greenhouse / Greenhouse area (NIS / square meter)	4	-0.727	-3.6 **
Logsum coefficient	All	0.610	1.7 *

* Significant at 0.1 level

** Significant at 0.05 level

Total observations:	148
Log-likelihood (null model):	-205.17
Log-likelihood (constants only):	-197.72
Log-likelihood (final model):	-172.38
Rho-bar squared w.r.t. null model:	0.16
Rho-bar squared w.r.t. constants:	0.13

The results presented in Table 3 above do not include alternative-specific constants, which were not significantly different from zero. However, this third model has additional dummy variables compared to the previous ones, and every dummy variable in the model reduces the significance of the constant.

All the variables in the table poses the expected signs, apart from the variable Ratio Investment in Greenhouse divided by Greenhouse Area. We expect this variable to be positive, indicating that greater investments would impact the selection of new varieties. This is partly explained by the time lag between the investment and the actual production, which cannot be captured by the static structure of discrete choice models used in this study.

Influence of R&D Center

This section presents selected results to illustrate the influence of a variable related to the R&D center. The distance of the agricultural settlement from the R&D center was selected for this purpose. This variable was found significant in Models 1 (choice of tomato varieties) and 3 (choice of tomato varieties conditioned on the choice of fertilization techniques). To compute the probabilities of growing a tomato variety, the following values were set for each of the remaining variables of the models:

- Percent export: 50%
- Greenhouse Area Ratio: 0.02
- Propensity to innovate: yes
- Use of new techniques: yes
- Awareness of R&D center: yes
- Visit R&D center: yes
- Settlement Age: 20 years
- Ratio Investment to Greenhouse Area: 5 NIS / square meter

Figures 1 and 2 show the results for Models 1 and 3, respectively. It can be seen that the probability to grow a new variety in both models is small. This is explained by the effect of two variables: the relatively high percentage export variable, which explains the fact that farmers prefer to grow known varieties, and the settlement age, which is relatively old for the region.

Both models show a similar pattern with respect to the old and current variety. However, in Model 1 the influence of the R&D center is more pronounced, which is explained by the magnitude of the coefficient in this model (-0.069) compared to Model 3 (-0.048). (Appendix, Figures 1 and 2)

Summary and Conclusions

This paper presents a new approach to model the adoption of innovations in the agricultural sector. The approach is based on discrete choice analysis, in which the choice sets are selected among different tomato varieties and fertilization techniques. In both sets, new techniques and varieties represent an alternative in the choice process.

Results of applying the disaggregated Logit Model indicate that regional, local and individual attributes have bearing on the farmers' decision-making process regarding the choice of tomato varieties. Among the variables found to be significant to the growers were percent export, the grower's use of advanced techniques, the age of the settlement, grower's relative share of greenhouses area in the settlement, distance form the regional R&D center, the inclination towards adoption of innovation, grower's age and the investments in greenhouses relative to the total area allocated to greenhouses in the settlement.

Regarding the decision processes of choice between various alternative fertilization techniques it was found that the most significant factors are: the age of the settlement, farmer's tendency to consult R&D personnel, growing new varieties, farmer's relative greenhouse area in the settlement, frequent visit to the regional R&D center and the extent of investments in greenhouses.

The paper presented also results for a combined model of adoption of alternative choices of fertilization techniques and tomato varieties. The advantage of this model with respect to the simpler ones is that two innovation processes are included in a joint structure, whereas the simpler models can only take into account a single innovation process. The fact that more than one innovation process is modeled raises interesting questions, such as the precedence of selecting one innovation process over the other. The results presented in this paper indicated that the choice of tomato varieties is conditioned by the choice of fertilization techniques.

The application of disaggregate choice models is conditioned by the quality of data available. This research collected extensive data on individual farmers, which enabled model specification and estimation. More research is needed to compare the applicability of such models to other issues that concern agricultural sectors.

References

Arndt, T. M., Dalrymple, D.G., and Ruttan, V. W. (1977). *Resource Allocation and Productivity in National and International Agriculture Research*. University of Minnesota Press, Minneapolis.

Ben Akiva, M. and Lerman S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Ma.

Cohen (Kedmon) M. (1997). Spatial Diffusion of Innovations in Agriculture and Regional Development: The Case of Greenhouse Tomatoes in Israel. Ph.D. dissertation, Technion, Haifa.

Davelaar, E. J. (1991). *Regional Economic Analysis of Innovation and Incubation*. Aldershot, Avebury.

Evenson, R. E. & Kislev, Y. (1975). *Agricultural Research and Productivity*. Yale University Press, New Haven and London.

Frenkel, A., Shefer, D. and Stephen, R. (2003). Public Policy, Locational Choice and the Innovation Capability of High-tech firms: A Comparison between Israel and Ireland. *Papers in Regional Science*, **82**, pp. 203-221.

Griliches, Z. (1957). Hibrid Corn: En Exploration in the Economics of Technological Change. *Econometrica* **25**(4), pp. 501-522.

Kamien, M. I. And Schwartz, N. L. (1982). *Market Structure and Innovation*. Cambridge University Press, Cambridge.

Mansfield, E. (1968). *The Economics of Technological Change*. Norton, New York.

Meir, A. (1981). Innovation Diffusion and Regional Economic Development: The Spatial Diffusion of Automobiles in Ohio. *Regional Studies*, **15**, pp. 111-122.

Metcalf, J. S. (1981). Impulses and Diffusion in the Study of Technical Change. *Futures*, **13** (5), pp. 347-359.

Rogers, M. E. (1983). *Diffusion of Innovations*. The Free Press, N.Y.

Staudt, T. A. and Taylor, D. A. (1965). *Marketing. A Managerial Approach*. Irwin, Homewood, Illinois.

Yeats, M. (1974). *An Introduction to Quantitative Analysis in Human Geography*. Mac-Graw Hill Book Company, New York.

Appendix

Figure 1. Influence of the R&D center (Model 1)

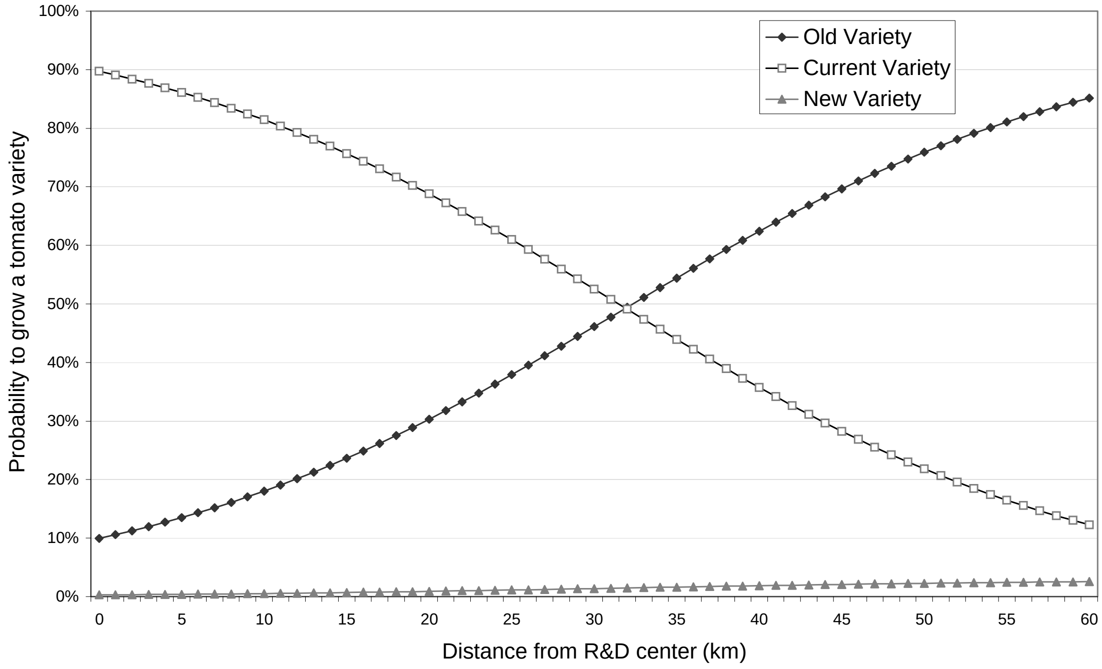


Figure 2. Influence of the R&D center (Model 3)