1975

The Urban Area Production Function and the Urban Hierarchy: The Case of Saskatchewan

Gordon P. Schaefer

Follow this and additional works at: https://ir.lib.uwo.ca/economicsipsle_rs

Part of the Economics Commons

Citation of this paper:
RESEARCH PROGRAM:
IMPACT OF THE PUBLIC SECTOR
ON LOCAL ECONOMIES

Research Study 02

THE URBAN AREA PRODUCTION FUNCTION
AND THE URBAN HIERARCHY: THE
CASE OF SASKATCHEWAN

Gordon P. Schaefer

October, 1975

Department of Economics
The University of Western Ontario
London Ontario Canada
Acknowledgements

This study is a condensed version of the author's Ph.D. dissertation. Copies of the original dissertation can be obtained from University Microfilms. Except for the expense of publishing this manuscript, the research was supported by a Canada Council Doctoral Fellowship while the University of Western Ontario supplied both on- and off-campus computer support. The study is published by this research program because of its relevance to work which may be done later in the program. The research program itself is currently funded principally by a grant from the Academic Development Fund of the University of Western Ontario.

The completion of this study owes a great deal to the conscientious and thorough criticisms of my thesis committee, Professors T. J. Courchene, D. F. Burgess, and E. F. Haites. Professors A. Berry and M. Frankena also gave considerable help in the study's formative stages. As well, F. Showell, K. Mackie, K. Love, and many other civil servants in the Province of Saskatchewan rendered invaluable assistance in acquiring essential data. Finally, the income tax data could not have been obtained without the programming assistance of David and Linda Pratt. As usual, all remaining errors are the responsibility of the author.

Gordon P. Schaefer
Post-Doctoral Fellow
Katholieke Universiteit te Leuven
THE URBAN AREA PRODUCTION FUNCTION
AND THE URBAN HIERARCHY: THE
CASE OF SASKATCHEWAN

Contents

I. INTRODUCTION ................................................. 1

II. THE ISSUES AND THEIR SOLUTION .......................... 6

III. THE MODEL .................................................. 35

IV. THE DATA .................................................... 53

V. CONSTRUCTION OF THE DISCRETE HIERARCHY .............. 73

VI. ESTIMATION OF THE MODEL ................................. 83

VII. ESTIMATION OF THE COST SHARE EQUATIONS .............. 107

VIII. SUMMARY AND CONCLUSIONS ............................... 114

APPENDIX I - SEPARABILITY CONSTRAINTS ...................... 118

APPENDIX II - A POPULATION HIERARCHY VARIABLE .......... 123

BIBLIOGRAPHY .................................................. 125
Chapter I

INTRODUCTION

The purpose of this study is to test various economic hypotheses about urban areas. This is done by estimating an urban area production function with cross-section data from 1971 for 464 incorporated communities in Saskatchewan. This study represents the first attempt at estimating an urban area production function and, as a result, is an important step towards implementing and evaluating economic models of the urban area. This approach facilitates the testing of hypotheses on returns to scale and income distribution. Finally, the thesis integrates the urban hierarchy into a model with an explicit urban area production function.

At the present time there are two distinct approaches to the urban area. The first views the urban centre as a phenomenon independent of external factors. The urban area's growth, size, and output are all determined by factors internal to the urban area. In part, this approach is adopted to keep the model simple and the question is whether too much has been sacrificed for the sake of simplicity. The second view, while accepting the general approach of the first, places the emphasis on external factors (e.g., competition from other centres, demand conditions, specialization of urban functions) and argues that these external phenomena are the crucial influences on the urban area. It is also argued that these external factors are reflected in the urban area's position in the urban hierarchy.

Specifically, this study tests the assumptions of several urban models to determine if they are consistent with empirical observation. The assumption common to most urban models is that scale economies exist in urban areas. While both approaches accept this postulate, they differ
as to the source of the scale economies (internal vs. external factors). By explicitly incorporating a hierarchy variable into the production function model it is possible to disentangle the two possible causes of returns to scale. As well, the estimation of the production function allows tests of input separability and of hypotheses on income distribution in a regime of non-constant returns to scale.

The urban area's output (value added) is measured by the sum of factor payments which is produced by aggregated inputs of labour, land, and capital. Real measures are used for labour (work force) and land (acres) while a market value measure is used to measure capital (assessments). The Transcendental Logarithmic (translog) Production Function is used to estimate the urban area production function. The vast majority of the 464 incorporated urban areas in Saskatchewan have a population of less than 50,000. This is in contrast with previous work which has, understandably, concentrated on much larger centres. However, the production function approach is perfectly general and the results and conclusions should be independent of the size range of the observations.

The policy implications of this study are most directly related to urban growth. By determining the extent of returns to scale in urban areas, and their source, the questions of urban viability and minimum size thresholds for growth are addressed directly. While the suggested sizes may be a function of the study area, the methodology used can be applied to any regional urban system. Of course, the estimation of the urban production function supplies half of the information required to determine the optimum size of an urban area—a question of some importance for urban planners. Estimation of the cost function would supply the remaining information. The degree of returns to scale is also related to the migration
issue since communities with relatively large scale economies will be more attractive to industry and migrants than other communities. Finally, the location of viable communities and potential growth centres is essential for any government development plan and permits the use of existing economic tendencies to implement policy objectives.

Two types of hierarchies, the continuous and the discrete, are statistically significant. It is found that returns to scale are variable. The smallest communities, which are also at the base of the hierarchy, exhibit decreasing returns. As urban size increases the returns to scale also increase and the largest communities exhibit increasing returns to scale. When non-constant returns to scale prevail at least one factor cannot receive its marginal product. In Saskatchewan, over the entire sample, there appears to be a transfer from capital to labour in the sense that capital receives less than its marginal product and labour receives more. However, for the largest centres this pattern of income distribution is reversed. Despite this last finding, labour seems to be the greatest beneficiary of urbanization. It is concluded that, while both the internal and external approaches to the urban area have a general descriptive validity, they are inadequate by themselves. Together they provide a relatively complete description of urban phenomena in Saskatchewan.

In Chapter II economic models of the urban area are outlined. The emphasis is on determining what economic aspects are included and which are excluded from the models. The standard approach postulates an urban area production function with increasing returns to factors within the urban area. External factors are ignored. An alternate view, that external factors matter, is considered under the heading of models of the urban hierarchy. In Chapter III the two approaches are synthesized and formulated
so that the hypotheses of the models are testable. Chapter IV discusses the sources, reliability, and possible biases of the data. Chapter V is a necessary departure from the main argument of the paper. To fully test particular versions of the urban hierarchy it is necessary to construct a "discrete hierarchy," and this occupies all of Chapter V. Chapter VI discusses the results of the direct production function estimation while Chapter VII deals with the cost-share estimation. The latter chapter is primarily concerned with questions of income distribution. Finally, Chapter VIII contains some concluding remarks and a summary of the paper.

For future reference and for purposes of locating the study area, a map of Saskatchewan is presented in Figure 1.
Chapter II

THE ISSUES AND THEIR SOLUTION

Introduction

This chapter discusses models of the urban area and of the urban hierarchy. The purpose of the discussion is to outline the models upon which this study is based, to indicate the assumptions which are testable hypotheses, and to consider possible modifications to the models when they appear inadequate. Empirical work bearing on these issues is also cited.

The assumption common to the urban models discussed is that scale economies exist in urban areas. Essentially this study is concerned with testing the validity of this assumption and with exploring the ramifications of its acceptance or rejection. Accordingly, for this study the important features of urban models are (1) the postulated rationale for scale economies and (2) how scale economies are incorporated into the model. At the risk of oversimplification, the issue is whether the source of scale economies is internal or external to the urban area.

Part I of the chapter deals with some of the existing models of the urban area. These models consider how the equilibrium size of the city is determined and what the characteristics of the optimum city are. Scale economies are assumed to be caused by the existence of extensive positive externalities in production and consumption. Negative externalities and decreasing returns to infrastructure and the public sector keep the urban area's size finite and raise the question of the characteristics of an optimum city. Both the positive and negative externalities mentioned in this connection are usually internal to the urban area. Since not all studies deal with these externalities (agglomeration economies) in any
detail a brief discussion of them follows the introduction to Part I.

Part II of the chapter shifts its attention to the urban hierarchy, i.e., the size distribution of urban centres in a region. The urban hierarchy is introduced because (1) it is one possible way of resolving some of the difficulties with the models discussed in Part I and (2) it is part of a debate about whether or not the urban area can be analyzed without reference to surrounding urban areas. The central issue seems to be whether the causes of urban growth are internal or external to the individual urban area. The proponents of the hierarchy argue that external factors are important, perhaps crucial, and could vitiate any results obtained from a model based on internal aspects alone.

Part II begins with an initial definition of the hierarchy which is followed by a brief discussion of the distributions which can describe the empirically observed distribution of city sizes. Alternative explanations of how these distributions can be generated are outlined. Each explanation has a different implication as to how the hierarchy should be attached to a theoretical model of urban areas using a production function. Part II concludes with two sections outlining empirical work on the hierarchy. The last of these sections deals specifically with work done on Saskatchewan.

I. THE EXISTENCE AND SIZE OF URBAN AREAS

There are several possible explanations for the existence of an urban area. Geographical factors (e.g., ore bodies, rivers, harbours), political factors (e.g., capital cities), sociological factors (e.g., religious centres), and other non-economic factors are all possible explanations. However, the
main economic justification for urban areas is that increasing returns to scale are associated with spatial concentration. A finite-sized urban area can be generated by assuming increasing returns to scale in the production of "gross" urban goods and decreasing returns to scale in the production of transportation services or public goods, or both.

In general, the equilibrium size and output of the urban area is determined by input supplies and the demand for output. The output consists of real goods and services which are either exported to the surrounding area or consumed internally. The models discussed below assume that one composite good may be constructed and treated as the urban area's output. They also assume that the size of the urban area is constrained by internal supply conditions. The supply of urban area goods is determined by the production function, transportation costs, and factor supply conditions with the production function being the crucial element. If locational factors are ignored then the extent of increasing returns to scale becomes the major determinant of the size and growth of an urban area.

Increasing returns to scale in the urban area can be attributed to economies at the firm or industry level or to agglomeration economies. The

---

1Mills (1972, p. 14) shows that positive and increasing transportation costs are not a sufficient condition for the existence of cities. He concludes that non-homogeneous land and/or increasing returns to scale are sufficient to justify the existence of cities.

2See Mills (1967, p. 204). If transportation is supplied at constant returns to scale then decreasing returns result from the longer trips required as the urban area's size increases, see Dixit (1973).

3To keep the city's size finite decreasing returns are required, and must be sufficient to offset, at some point, the increasing returns. With increasing returns to urban goods the decreasing returns must be found in transportation services or public goods or both.
latter term refers to the advantages of spatial concentration that result from the scale of an entire urban area rather than from the scale of a particular firm or industry.\textsuperscript{4}

A. Agglomeration Economies

Agglomeration economies or diseconomies refer to positive and negative externalities in production and consumption. Accordingly, they can be divided into two groups,\textsuperscript{5} those accruing to households and those accruing to firms. The benefits to households of increased urban size include the wider range of available services, e.g., entertainment, food, cultural activities, and medical and educational facilities. As well, the set of market and employment opportunities for workers is larger, while access to information about new or other jobs is likely better.

The possible benefits to firms are more numerous, but, in the main, result from inter-industry production externalities. External economies may result from the ability to share services and facilities, thus lowering costs. These cost savings result from indivisibilities in the provision of, for example, public utilities such as electricity and sewage disposal. Transportation savings and access to specialized business services which require a minimum volume to function and can only exist in large concentrated markets, e.g., advertising, lawyers, repair shops, rental markets, may also lower costs. In some industries proximity to other firms or industries may lower costs if it facilitates face-to-face communication, allows easier

\textsuperscript{4}Mills (1972, pp. 16 and ff.).

\textsuperscript{5}See Marcus (1965), Isard (1956), Nourse (1968), Conrod (1973), Mills (1972), Chinitz (1961), and Vernon (1960).
comparative shopping, and if the alternative to a city with its stock of goods and services is the carrying of a large private inventory of inputs.

Diseconomies also exist. Higher factor prices, e.g., wages and rent, relative to those in smaller centres would represent diseconomies to the firm to the extent that the increases are not due to higher productivity. Other examples of diseconomies are crime, the hectic life, pollution and congestion.\(^6\) Crime and pollution may have adverse affects on productivity and they require real resource expenditures to control them, e.g., waste disposal systems and law enforcement. Congestion costs refer mainly to increasing per capita transportation costs. Some of these costs will be capitalized in land values, but some will be reflected in higher wages to workers as compensation for the inconveniences.\(^7,8\)

In sum, agglomeration economies result from indivisibilities and inter-industry production externalities but they are not produced simply by the urban area's size: the number, type, and size of different activities are also important. As used in the models of Alonso, Edel, and Conrod, the size of the urban area becomes a proxy for all of the household and firm externalities.

\(^6\)Hoch (1972, pp. 308-24) discusses some empirical evidence on these economies. With regard to crime he finds positive and significant coefficients for population when robbery and auto thefts are the dependent variables. Also, population density is positively related to robbery, auto theft, assault, and burglary rates.

\(^7\)Hoch (1972), Edel (1971), Evans (1972), Mera (1973), and Alonso (1971).

\(^8\)In part, the urban area's size is finite because of such diseconomies. Other reasons include the demand for the urban area's product and limited access to basic resources such as water.
B. The Alonso-Edel Model

Alonso (1971) and Edel (1971) have discussed a simple model of urban size based on increasing returns. Except for the income distribution assumption and for minor differences in terminology, their models are identical. This model is important because (i) it treats the city as an aggregate productive unit, (ii) it defines output as the value of the total product of the urban area which is equal to total factor payments, and (iii) it indicates a possible approach to the income distribution problem when increasing returns to scale prevail. This model is the simplest version of all the production function models and is the base upon which this study builds.

Alonso takes urban output as the value of the total product of the urban area. This total product function yields the average product (AP) and marginal product (MP) curves in Figure 2. The upward slope of the MP curve embodies the increasing returns assumption. The population of the urban area serves as a proxy for all the inputs of the urban area. He is implicitly postulating an urban area production function. To keep the city's size finite the costs of the urban area must be included in the model. However, these costs remain vague:

Urban costs are harder to define, and would include quantity and price effects in the costs of infrastructure and municipal operation, in the costs of exogenous inputs other than human ones into the city's economic activity and in private consumption.9

Edel, in discussing the use of land values as a measure of the net costs or benefits of agglomeration, employs a model like Alonso's. His

---

9Alonso (1971, p. 70).
Figure 2.

From Alonso (1971, p. 71)
average cost curve (AC) includes the fiscal burden of providing public services through taxation, commuting costs, and the physical costs due to noise and air pollution, crime, etc. He uses an average benefit curve (AB) based on the agglomeration economies discussed in the previous section. Both the total output of Alonso and the total benefit function of Edel exhibit increasing returns. As well, both authors assume that the net benefit or output of the urban area exhibits increasing returns over some range of urban population.

As the above quote suggests, Alonso does not include labour costs in his cost curves. The "return to labour (in the broad sense of total urban population) is the difference between the value of total output and total costs." Of course, this does require that the opportunity costs of labour outside the city be included in the cost curves. Edel and Harris and Wheeler (1971) include labour costs and suggest that the difference between output and costs goes to land. This is the one point on which Alonso and Edel diverge. In either case, because of increasing returns at least one factor cannot be paid its social marginal product. For both Alonso and Edel only one input receives less than its social marginal product, but of course this could be the case for any number of factors.

What happens in this model depends on how income payments and decisions are made in the urban area. If each worker is self-employed then the difference between the average product and the average cost curves

---

10 Edel is one of the few authors to include consumption externalities explicitly.

11 Alonso (1971, p. 70).
measures his return. If perfect competition prevails then the city grows until AC = AP, i.e., until $P_e$ is reached (see Figure 1). A planner with a national outlook would stop the city's growth at $P_o$ where the city's contribution to national output is maximized, i.e., $MC = MP$. A city planner might choose to keep the population at $P_b$ where $(AP - AC)$ is maximized, i.e., $(dAP/dP = dAC/dP)$. Depending on the objective, either $P_o$ or $P_b$ could be considered the social optimum versus the "private" equilibrium at $P_e$.

The unstated assumptions underlying this simple model are that (i) there is only one city or a fixed number of cities and (ii) relative prices do not change with urban size so that a composite urban good can be constructed. Increasing returns and assumption (i) are the crucial ones required to obtain the static results discussed above.

The model can be summarized as follows:

1. $Z = F(P)$, $F' > 0$, $F'' > 0$,
2. $C = G(P)$, $G' > 0$, $G'' > 0$,
3. Alonso: $Z - C =$ total wages in excess of opportunity cost
   Edel: $Z - C =$ total rent.

where $Z$ is total output, $C$ is total cost, and $P$ is population.\(^{12}\) All these points, including increasing returns, are testable hypotheses. The estimation of the production function for gross urban goods provides sufficient information to evaluate all these points save for (2), the cost function. The assumptions embodied in (1), (3), and the variations of them in other models, are what this study seeks to test empirically.

\(^{12}\)To guarantee a finite city size greater than zero requires that $F'(0) > G'(0)$ and $G'' > F''$ for all $P \geq P$, $P > 0$. 
C. Models with Explicit Production Functions

(i) Introduction

This section considers three models of the urban area in detail. The models of Mills (1967) and Dixit (1973) use Cobb-Douglas production functions with increasing returns due to agglomeration economies. Both these models make assumptions while rule out external influences on the city thus making the urban area self-contained. Conrod (1973) assumes variable returns to scale and also confines himself to internal factors. His approach is discussed first since it is simpler and indicates how external factors might be introduced into the model. However, Conrod does not discuss the income distribution problem whereas Mills and Dixit recognize the need for special assumptions to resolve it.

(ii) Conrod

Conrod uses an explicit production function but retains a city size variable. The latter affects factor productivity and serves as a proxy for increased factor specialization and agglomeration economies. Thus, he writes \( Z = A(p) \cdot F(K, L) \) where \( Z \) is the aggregate output of the city, \( K \) is the service flow from capital, \( L \) is the service flow from labour, and \( A(p) \) is a shift parameter which augments factor productivity and is a function of city size \( (p) \). To keep the city's size finite he assumes \( A''(p) < 0 \).

While Conrod seems to use only internal aspects of the urban area to justify his results, this is not necessary. Unlike the Alonso-Edel model, or the ones which follow, Conrod's city need not be self-contained. His model does not rule out other urban areas nor is such an assumption required. It would be necessary, however, to add another equation which determines \( p \). To anticipate what follows, this can be done via the urban
hierarchy with a central place model.

(iii) Mills

Mills (1967) uses a three input Cobb-Douglas production function,

\[ Z = A_p E_p^{\alpha_1} L_p^{\beta_1} K_p^{\gamma_1} \; ; \; \alpha_1 + \beta_1 + \gamma_1 \leq 1 \]

Z is goods production and \( E_p \), \( L_p \), and \( K_p \) are, respectively, the land, labour, and capital used in goods production. If there are no locational advantages to the city's site then \( \alpha_1 + \beta_1 + \gamma_1 > 1 \). Goods production is located in the central business district (CBD). Because he is interested in the division of land into alternative uses he adds a housing sector which has a three input Cobb-Douglas production function with constant returns to scale.\(^{13}\)

Increasing returns to scale are assumed and are due, again, to internal agglomeration economies. The production function could be modified to handle external factors by making either \( A_p \) or the efficiency parameters a function of, say, city size. The present study does investigate these possibilities but such a procedure makes the solution to Mills' questions very complicated.

Because of increasing returns the goods producer must be assumed to be a monopolist. Factor markets are assumed to be competitive so that all activities pay the same factor prices. The wage rate, \( w \), and the rental rate on capital, \( i \), are assumed to be exogenous.\(^{14}\) Mills suggests that these

\(^{13}\)Mills (1967, p. 202): "The assumption is that all commodities whose production functions have non-constant returns to scale and whose efficiency parameters are affected by location can be aggregated into the production function for goods. Competition will force the production of all other commodities to be located adjacent to customers in order to avoid transportation costs." This latter aggregate is called housing.

\(^{14}\)It must also be assumed that the city is small relative to the rest of the economy.
are the appropriate assumptions if the city's size is to be endogenous.

Increasing returns create difficulties in handling income distribution in the urban area. Mills' solution is to pay labour and capital their marginal products. Total rent paid in the CBD absorbs any monopoly profit, i.e., \( rE_p = pZ - wL_p - iK_p \), and land is the only factor which does not receive its marginal product. This model can be solved for the rent function, the population density function, and the equilibrium land allocation between goods and housing.

(iv) Dixit\(^{15}\)

Dixit's production function is \( Z = AL_p^\alpha E_p^\beta \) where \( E_p = \pi c^2 \) and \( c \) is the radius of the CBD. He assumes \( 0 < (\alpha, \beta) < 1 \) and \( \alpha + \beta > 1 \), i.e., decreasing returns to each factor and increasing returns to scale. Part of \( Z \) is used to pay landowners and a portion \( Q = aE \) is an input for production which can "... be thought of as implicit rental costs of plant and equipment, or as public services."\(^{16}\) The remainder of \( Z \) is allocated to households according to the distance they live from the CBD. All production is distributed to factors as income. Again land receives the residual after labour is paid its marginal product. He assumes that the population, \( P \), is to be divided into self-sufficient towns each of size \( N \). As in the case of Mills, any attempt to introduce external phenomena will make the model very cumbersome and difficult to solve.

\(^{15}\) Dixit's model is part of the optimum city literature [Riley (1973), Mirrlees (1972), Oron et al. (1973), Dixit (1973)], which is concerned with determining the form and size of the urban area which will maximize a given social welfare function. At some point a production function must be introduced. Riley and Oron et al. assume constant returns to scale but Mirrlees and Dixit explicitly postulate increasing returns to scale.

\(^{16}\) Dixit (1973, p. 639).
(v) Implications and Summary

All four models assume increasing returns to scale for an urban area production function. Generally these returns are due to agglomeration economies which are internal to the urban area. To keep the city's size finite all the models use factors internal to the city. Except for Conrod's model they all assume decreasing returns to the production of transportation services and public goods, i.e., increasing congestion costs. While the production function for gross output ($Z$) exhibits increasing returns to scale throughout its entire range, net output ($Z$ minus congestion costs) does not. Conrod simply lets $Z$ cease to exhibit increasing returns after a certain size is reached. All external factors, such as the demand for the urban area's output and competition from other centres, are ignored or assumed to be non-binding constraints. One of the questions this study seeks to answer is whether this is a reasonable and empirically viable approach.

Part of the reason why so much emphasis has been placed on external-internal factors is that the models, as outlined above, cannot produce optimum cities much in excess of 1,000,000 people.¹⁷ This result implies that either the real world deviates considerably from optimality or that the model is omitting some important factors. There are at least two ways to modify the model; either production can be decentralized or external factors such as the urban hierarchy can be introduced. The first approach could reduce congestion costs and increase the optimum city size; the

---

¹⁷This conclusion is based on Dixit's simulations of the optimum town. Dixit (1973, pp. 645-650). The welfare function which is maximized is a constant elasticity utility function (p. 640).
second introduces the possibility of different urban areas serving different purposes and having different sizes. As well, the introduction of unpaid external factors could allow a simple resolution of the distribution problem when increasing returns to scale prevail.

D. Empirical Evidence on Agglomeration Economies

The empirical work on agglomeration economies and city size is limited. Edel (1971) and Harris and Wheeler (1971), using the Alonso-Edel model as their basis, attempt to measure the net benefits of an urban area and to determine the size at which these benefits are maximized. Further information on the average product and cost curves of urban areas is discussed by Mera (1973) in a survey of the empirical work on social overhead cost and productivity in urban centres.

Mera assumes that the objective is to maximize national income and attempts to demonstrate that even the largest metropolitan area in the world is likely to be below the "optimal" size. He first considers how the costs of operating an urban area vary with population density; he then discusses the relationship of factor productivity and urban size. He argues that "... the expenditure per capita is not very much related to the population size when the expenditure is one way or another adjusted for quality differences."18 For Japan he notes that there are substantial economies of higher density in social overhead capital stock. He concludes that "optimal" size is more likely to be determined by productivity per capita than costs per capita. His analysis suggests that there are either

---

18Mera (1973, p. 312).
constant or increasing returns to infrastructure but that these are not
crucial. The most important factors are productivity conditions and it
appears that the economies of concentration outweigh the diseconomies. No
reference is made to external factors.

Harris and Wheeler argue that total land value represents a lower
bound measure of the total net benefit \((Z - C)\) associated with any urban
area.\(^{19}\) To test their model they run regressions on population, \(P, P^2,\) and
\(P^3\) as well as other environmental variables, e.g., universities, transporta-
tion index, mineral value index, etc. The population variables are supposed
to measure the urban area's size and their coefficients indicate whether
net benefits are increasing or decreasing. The other variables are used to
control for differences in the activity mix between cities.\(^{20}\) They find
that, for metropolitan areas with more than 50,000 inhabitants in 1960,
total land value rose with city size. In loglinear multiple regression
analysis the elasticity of total land value with respect to population was
approximately unity. This would seem to indicate that constant returns to
net output exist which implies either constant or increasing returns to
gross output, depending on the returns to the cost function.

Edel is also concerned with measuring the net benefits of an urban
area. He argues that land prices are an imperfect measure of the net
social cost of congestion. One problem with land values as a measure of net

\(^{19}\) Dixit's simulations suggests that total income or wage payments
will overstate total benefits and indicate a larger optimum. It is possible
that these suggested biases could be used to measure net benefits and to
determine optimum sizes more accurately.

\(^{20}\) This is suggestive of the urban hierarchy, see below.
benefits is that higher factor payments can serve as compensation for some congestion costs. These compensation payments would cause land values to overstate total net benefits. The second major objection is that different urban areas may have different benefit functions because of the different roles they serve. Edel suggests three general categories of cities and that there may be an optimum size within each category. From his empirical work he concludes:

... on the average, economies of agglomeration outweigh costs associated with agglomeration for cities of a size up to at least half a million in population. For metropolitan areas of somewhat larger size the balance is more doubtful. For some larger cities, ..., rising land values again indicate a private net benefit to those taking part in urban agglomeration. However, these benefits seem limited to cities with certain corporate headquarters functions. ... the possibility emerges that congestion costs are being incurred for the maintenance of the hierarchy.

In conclusion, increasing returns to scale seem to exist. However, the results are not directly applicable to the production function. Finally, if Edel's suggestion is correct, then the postulated internal causes of agglomeration economies may not be crucial. It could well be that external factors, as represented by the urban hierarchy, supply the

---

21 Hoch (1972) argues that this is a significant factor.

22 Edel (1971, p. 19). He calls them non-corporate, corporate, and world cities. In effect he is postulating three different production functions.

23 Richardson (1972, p. 35) also suggests that if a well-defined hierarchical level exists, then "... instead of a single optimal city size it may be more meaningful to conceive of an optimal size for each rank-order in the urban hierarchy."

24 Edel (1971, p. 8).
main impetus for growth and are the major determinants of urban size. Hopefully this study will shed some light on this issue.

II. THE URBAN HIERARCHY

A. (i) Introduction

The theoretical models of Part I confined themselves to one urban area or to a finite number of equal-sized urban areas. It was suggested that if these models were to serve as a framework for empirical work, they would have to be modified to include the urban hierarchy, i.e., "... (the) broad base of many smaller size urban centres underlying a smaller number of larger cities ranging up to a dominant regional metropolis."\(^{25}\) This Part of the Chapter considers the empirical characteristics of the urban hierarchy in general and then proceeds to discuss possible explanations of the hierarchy which are consistent with empirical observation. The rest of section A discusses city-size distributions while section B considers models which can generate one of the acceptable distributions. The implications each model has for the urban area production function are indicated and the theoretical and empirical problems of each model are discussed. Section C discusses empirical work using some of these models while section D discusses some empirical work on Saskatchewan. A final section summarizes the material of Part II.

(ii) Introduction: City Size Distributions

All city size distributions are positively skewed to the right, i.e., there are many small but only a few large cities with a tendency for the

number of cities in each size class to decline as city size increases. The lognormal, Pareto, and rank-size distributions are all consistent with observed urban size distributions. Accordingly, a model of the urban hierarchy must be able to generate one of these distributions.

The three distributions mentioned in the preceding paragraph are all similar through at least part of their range; that is why it is so difficult to accept one or the other as the superior representation. The rank-size distribution can be written as

\[(1) \quad R^q p_R = C \]

or,

\[(2) \quad \ln(p_R) = \ln C - q \ln R \]

where \( R \) = the rank of a city; \( p_R \) = the population of a city of rank \( R \); \( C \) = the population of the largest city; and \( C \) and \( q \) are constants. The rank-size rule, \( q = 1.0 \), is a special case of the rank-size distribution. The Pareto Distribution is written as

\[(3) \quad N = A/x^b \]

or,

\[(4) \quad \ln N = \ln A - b \ln(x) \]

where \( N \) is the number of cities with a population greater than \( x \) and \( A \) and \( b \) are constants. The difference between the rank-size and Pareto distribution lies in the variables used, not in the functional form.\(^{26}\) The lognormal

\(^{26}\)Parr (1970, p. 238) and Richardson (1973).
distribution can be written as

\[ \ln(x) = \sum_{i=1}^{n} w_i \ln(x_i) \]

where \( x \) is the population size of the urban area and the \( x_i \) are well-behaved random variables which affect \( x \). The lognormal distribution can also be written as \( N = \ln P \) where \( N \) = cumulative percentage of cities and \( P \) = city size. The lognormal attempts to represent the whole range of city sizes whereas the Pareto distribution deals only with the upper tail (city sizes above a defined level, \( P \)). However, the lognormal can be translated into Pareto distribution terms by imposing a threshold city size.\(^{27}\) (See Figure 3a.) Figure 3b compares the lognormal and rank-size distributions.

The three distributions are all non-hierarchical or continuous. Simply put, all urban areas are ordered by population size with the largest centre receiving the rank of one. Smaller centres are given the appropriate number ranking. In contrast a hierarchical or discrete distribution has a finite number of levels and within each level all centres would have the same population. This is illustrated in Figure 4.

B. Models of the Urban Hierarchy

(i) Central Place Models

One of the simplest models of the urban area is the central place theory.\(^{28}\) Scale economies, in the form of initial fixed production costs and constant marginal costs, are assumed. The point of these assumptions

\(^{27}\) Richardson (1973, p. 240).

\(^{28}\) Lösch (1954, Christaller (1966), Beckman (1958), Beckman and McPherson (1970), Henderson (1973), Berry and Garrison (1958a), and Mills (1972).
Figure 3

a) Number of Communities

Lognormal

--- Pareto
--- Lognormal

Size

Figure 4

A Hierarchical Distribution

Cumulative Frequency or Rank (in logs)

Size

P1 P2 P3 P4

Number of Communities
is to generate a downward sloping average cost curve. This serves as the rationale for the existence of the urban area. Transportation costs define the trading area of the urban centre.\textsuperscript{29} To this point all the communities produce the same good, are of the same size, and have the same market area. The hierarchy is produced by introducing more goods with different production and transportation costs.

Each method of introducing additional goods has a different effect on the predicted pattern of central places. Berry and Garrison (1958a) use one of the least restrictive approaches. Each good has different production and transportation costs and, therefore, different market areas. The good(s) with the smallest area(s) can be supplied by every community. The maximum number of such minimum areas a region can hold determines the maximum number of primary communities in the region. Goods with the next largest market areas are then produced in the primary communities with the most efficient locations. This generates the second level of the hierarchy and the process continues until all the goods are supplied. This procedure results in a discrete hierarchy, but as the number of goods increases the number of hierarchy levels will increase and, in the limit, a continuous hierarchy will result.

The Berry and Garrison approach is appealing and since it does not generate an \textit{a priori} population distribution it is compatible with all the continuous distributions. The drawback is that a community's population cannot be predicted. The classical model of Lösch and Christaller makes

\textsuperscript{29}Economies of scale can be associated with the concept of a demand threshold, the minimum viable level in population and/or income required to support a service. Transport costs are associated with the range of a good, the outer limit of the market area.
such predictions. 30 This is the type of formulation that would allow the
inclusion of external factors into the models of Part I. The question is
whether such a model is consistent with empirical observation.

The discrete hierarchical distribution produced by the classical
central place model is never observed and it is necessary to supply a
mechanism which generates one of the continuous distributions. Parr (1969,
1970) has convincingly argued that most central place models are inconsistent
with the rank-size rule. While the general validity of the rank-size rule
may be disputed, this remains a problem. However, the central place model
can be consistent with either a Pareto or lognormal distribution. 31
Accordingly, the central place model can be accepted as a possible model of
the urban hierarchy. 32,33

\[ P_m = \frac{kr}{s} \left( \frac{s}{1-k} \right)^m, \quad s > 1, \quad 0 < k < 1, \]

where \( P_m \) is the population of a place of the \( m^{th} \) order, \( k \) is the proportion of
total population served which is located in the central place, \( s \) is the number
of places of the \( m-1 \) order served by places of order \( m \), and \( r \) is the rural
population served by a first order place. Of course, the equation for \( P_m \)
is model specific.

31 Parr (1974). In any event, Parr and Suzuki (1973) indicate that the
lognormal distribution supplies a better empirical fit to data than a
rank-size distribution.

32 Another problem with the central place theory is that shopper surveys
indicate that, in going to market, shoppers skip hierarchy levels. [Johnson
and Rimmer (1967) and Stabler and Williams (1973)]. This is inconsistent
with a simple central place theory, though not with more complex ones,
[Parr (1973)], which allow multipurpose trips.

33 Variations on the central place model are possible, see Richardson
(1973) and Tinbergen (1968).
In the context of the present study a central place or discrete hierarchy model implies a different production function for each hierarchy level. Exactly how they differ is a moot point; efficiency parameters, functional form, and constant terms can all vary. Such a model also has implications for the type of returns that should be observed. Because firms will be monopolists within their market area increasing returns to scale should be observed.\textsuperscript{34} Decreasing returns are acceptable only in disequilibrium while constant returns would require highly competitive markets. Thus, constant returns are more likely to be observed in large urban centres, \textit{ceteris paribus}. However, \textit{ceteris paribus} will not be valid if industries with greater returns to scale also tend to locate in the larger centres. The suspicion is that increasing returns are the more likely result.

A central place or hierarchical model implies, in equilibrium, at least constant returns to scale, though the more likely result is increasing returns to scale. However, in these models the location of economic activities is determined by factors external to the urban area, i.e., locational factors generate growth and determine the extent of increasing returns within the urban area by affecting the firm and industry mix in a community. This is a phenomena distinct from the agglomeration economies which are internal to the urban area. Thus, the existence of increasing returns does not imply the existence of agglomeration economies. If, after controlling for the hierarchy, increasing returns to physical inputs still prevail, then agglomeration economies exist and the models of Part I would be supported. This is the central issue around which the internal-external debate revolves.

\textsuperscript{34}This result is analogous to that for any monopolistically competitive firm.
(ii) **Other Explanations**

Another explanation is given by Evans (1972). He uses the economic theory of clubs. He concentrates on the firm's location decision, but the model is essentially an extension of Tiebout (1956). Each city is treated as a coalition of firms with an associated set of costs and benefits. Firms locate so as to maximize profits and the city's size is a variable affecting the firm's cost function. This model can generate a discrete hierarchy and the connection between the firm and the urban area production function is explicit. Though there should be a tendency for constant returns to scale to be observed, the model requires more assumptions before detailed predictions can be made.

Henderson (1974) models city size by extending the Mills model of Part I. He assumes that cities specialize in different traded goods with different returns to scale. The greater the returns to scale the larger both the optimum and equilibrium size of the city. Accordingly, this model assumes, rather than predicts, that movements up the hierarchy are associated with a greater degree of returns to scale. This is a testable hypothesis. As the model makes no statement as to urban size distribution it can be considered compatible with continuous or discrete hierarchies.

A final approach is to use a stochastic model which treats urban growth as proportional to city size. Each urban area is assumed to have

---

35 In the Tiebout model individuals move between cities to associate with people having similar preferences for public goods and services and for tax-expenditure mix.

the same average growth prospect. But this average is subject to statistical variance; at any point in time different urban areas will grow at different rates. This type of process yields a lognormal distribution and has also been used to justify the observed size distribution of firms.\footnote{37} Stochastic models generate continuous hierarchies.

From a theoretical and practical standpoint the stochastic approach is not very useful since the causes of growth are placed outside the bounds of predictive theory. The approach is really saying that there is no theoretical or economic rationale underlying the urban hierarchy. This seems unsatisfactory, "explanations hanging upon unnamed 'forces' which are 'many and act randomly' are not theories at all but rationalizations for lack of a theory."\footnote{38} However, one advantage of the approach is its immediate generation of an acceptable urban size distribution. The stochastic "model" does not generate any predictions about the production function or the role of the hierarchy. While all the other models predict a positive role for the urban hierarchy, the stochastic approach alone would be consistent with a zero or negative influence.

C. Empirical Work

Most of the models of the urban hierarchy generate a discrete hierarchy and then use probabilistic mechanisms to derive a continuous hierarchy. The central place model is the most well-developed and most empirical work has

\footnote{37}If cities are added at a constant rate a Pareto Distribution results.\footnote{38}Higgs (1970, p. 253).
been based on it.\textsuperscript{39}

As was indicated above, either a continuous or a discrete hierarchy can be constructed. Aside from stochastic models the only empirical work on a continuous hierarchy is Woroby (1973). Using a set of characteristics describing each community as well as the distance between communities, he has been able to construct a continuous hierarchy for the area surrounding the city of Swift Current in Saskatchewan.

The construction of a discrete hierarchy is usually based on central place theory [Berry and Garrison (1958b, 1958c), Abiodun (1967, 1968), Davies (1967), Higgs (1970), Stabler and Williams (1973) and Berry (1961)]. The communities are grouped with factor and/or cluster analysis.\textsuperscript{40} Grouping characteristics are chosen \textit{a priori}. Because the selection of the characteristics is highly subjective the sensitivity of the resultant groupings to different sets of characteristics should be examined. Most studies do not include a sensitivity analysis, but do, in general, obtain reasonably well-defined groups of communities. The number of groups will vary from region to region. Berry and Garrison (1958b) found that three groups were adequate for Snohomish County, Washington, while Stabler and Williams had six groups for Saskatchewan's Qu'Appelle River Basin. The studies generally obtain similar rankings of central functions, e.g., dentists are invariably found to represent a higher order function than filling stations.

\textsuperscript{39}Aside, that is, from research directed towards determining which of the continuous distributions best describes the data.

\textsuperscript{40}See Harmon (1968).
One variable that is usually omitted in these studies is the distance between communities. One would expect that the range of functions offered by a small community would decline the closer that community is to a much larger centre. For verification one need only look at the "bedroom communities" attached to large metropolitan areas.

D. Empirical Work in Saskatchewan

This section contains just a small sample of the work done on Saskatchewan; it is, however, indicative of what has been done and the type of results which have been obtained.

Hodge (1965, 1966, 1968) and Stabler and Williams (1973) have carried out studies for Saskatchewan. Hodge uses factor analysis to group his communities and to isolate socio-economic aspects of viable and non-viable trading centres. The variables used for each community included population, ethnic composition, school and hospital quality, manufacturing and service labour force, agricultural characteristics, railway and highway access, and the distance to the nearest provincial centre and to the nearest competing centre. His main conclusion was that small communities offering a minimum number of services will continue to decline with the decline being greatest for communities closest to large trade centres. He argues that the volume of grain shipped from a community has little effect on its viability.

Stabler and Williams use cluster analysis to group communities in the Qu'Appelle River Basin. They obtain six groups, all significantly different from each other at a 99% confidence level, and examine the

41 Woroby (1973) and Hodge (1965) are notable exceptions.

42 Assuming that the groups represent independent random samples from
movements between these groups from 1961 to 1970. As is usual with such studies, there is some overlap between the functions found in each group. Figure 5 illustrates the situation in the Stabler and Williams study area in 1970.

Stabler and Williams conclude that the fourth and fifth groups are declining while the other groups are stable or growing. They suggest that changing shopping patterns, due in part to an upgrading of the road network, are largely responsible for the change.

E. Summary

At the conclusion of Part I it was indicated that the addition of the urban hierarchy to standard urban models was a possible method for resolving some of the difficulties of these models. Part II considered various aspects and models of the urban hierarchy. Of all the models discussed the central place model yields the most predictions about the production function. It predicts increasing returns within a hierarchy level and suggests that a production function should be estimated for each level. Henderson's model assumes that the degree of returns to scale increases as one moves up either a continuous or discrete hierarchy. Other models yield fewer testable hypotheses but do suggest that the hierarchy is important. The stochastic model is the sole exception. However, the central place model is the one which has been tested the most and it has borne the scrutiny well. Accordingly, much of this study is directed towards implementing a discrete hierarchy.

a multivariate normal universe Rao's F-test can be used to determine when the group's mean vectors are all significantly different. At that point we may stop clustering.
Figure 5. The Range of Functions Present in Each Class of Center, 1970.

SPECIALTY STORE (Florist, Music)
FURNITURE
DRY CLEANING
ACCESSORIES (Shoe Store, Jewellery Store)
LIQUOR BOARD STORE
VETERINARIAN
SPECIALTY CLOTHING
DENTIST
PRINTER
BAKERY
GENERAL CLOTHING
RCMP DETACHMENT
MEAT MARKET
TV–RADIO REPAIR
LAWYER
HOSPITAL
GENERAL CONTRACTOR
LOCKER PLANT
DOCTOR
DRUG STORE
AUTO REPAIR
AUTO SALES
BANK
LUMBER YARD
SPECIAL CONTRACTOR
PROV. GOVT OFFICE
RESTAURANT
HARDWARE
BULK FUEL DEALER
IMPLEMENT DEALER
HOTEL
GASOLINE STATION
SCHOOL
GROCERY STORE
GRAIN ELEVATOR

Generally, centers in each class will have all functions up to the solid line and may have some of those up to the dashed line. Width of the bar indicates the number of centers in each class. One inch is equal to 30 centers. A, B₁, B₂, C₁, C₂, and C₃ are hierarchy levels, with A being the highest level.

Taken from Stabler and Williams (1973, Figure 3)
Chapter III

THE MODEL

I. INTRODUCTION

This chapter considers the specification of the urban area production function, the role of the urban hierarchy in a model of urban areas, and how the hierarchy can be implemented with the urban production function. In the first section the general form of the production function is described. The next section deals with the functional form used to implement the model. Two forms of the estimating equation are discussed. A solution to the income distribution problem when non-constant returns to scale prevail is outlined. Also, this section outlines the technical restrictions on the production function and discusses various hypothesis tests. A third section considers the economic interpretation and role of the urban hierarchy in the model. The fourth section discusses possible forms of the estimating equations when the hierarchy is added to the model. A final section summarizes the chapter.

II. THE AGGREGATE URBAN AREA PRODUCTION FUNCTION

The assumption of increasing returns to scale in an urban area requires the existence of positive production externalities between firms and industries or increasing returns within individual sectors. If there are m industries or firms in an urban area the i\textsuperscript{th} industry's (firm's)

\footnote{Of course, the existence of increasing returns is a testable hypothesis. Without loss of generality it is used here as an expository device and for consistency with the models discussed in Chapter II. As well, household consumption externalities are not considered explicitly.}
output is a function of all other industry (firm) outputs as well as its
own inputs, i.e.,

\[(1) \quad z_i = f_i(k_i, l_i, e_i, m_i; z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_m)\]

for all \(i = 1, \ldots, m\) industries (firms). The output of the \(i^{th}\) industry (firm)
is \(z_i\) and \(k_i, l_i, e_i, m_i\) are the capital, labour, land, and material inputs
of the \(i^{th}\) industry (firm). Adopting the maintained hypothesis that the
production function is separable between material inputs and the primary
factors of land, labour, and capital implies the value added production
function

\[(2) \quad v_i = g_i(k_i, l_i, e_i; v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_m)\]

where \(v_i\) is the value added of the \(i^{th}\) industry (firm). If it were possible
to estimate equation (2) the parameters on \(k_i, l_i,\) and \(e_i\) would indicate the
extent of internal returns to scale while the parameters on the \(v_j\)'s would
represent the effects of agglomeration economies.\(^{44}\)

For the purposes of this study specification (2) is too detailed and
firms and industries must be aggregated. Let \(v\) and \(x\) be vectors of value
added and primary inputs in the urban area. Assuming that the transformation
function \(T(v, x) = 0\) is separable between inputs and outputs implies

\[H(v) - F(x) = 0.\]

This implies that a consistent aggregate index of the \(v_i\)'s exists, i.e.,

\(^{44}\) While equation (2) cannot be implemented at the firm level because
of data restrictions, it could be tested at the industry level. This is an
area for further research.
\[ V = \sum_{i}^{m} \frac{\partial H}{\partial v_i} (v_i), \text{ such that } V = F(x). \] 
The vector \( x \) consists of all the different types of capital, land, and labour used in the urban area. To reduce \( x \) to three aggregate inputs the maintained hypothesis of separability between all types of capital inputs, all land inputs, and all labour inputs is adopted,\(^{45}\) i.e.,

\[
(3) \quad V = F(K,L,E)
\]

where \( K \) is the flow of capital services, \( L \) is the flow of labour services, and \( E \) is the flow of land services for the entire urban area. Assuming constant utilization rates, service flows are proportional to the stocks of the inputs and stock measures can be used in equation (3).\(^{46,47}\)

III. THE FUNCTIONAL FORM OF THE PRODUCTION FUNCTION

A. The Translog Production Function

Estimating a production function poses two serious problems. The

---

\(^{45}\) Data restrictions, which are discussed in Chapter IV, make this assumption necessary.

\(^{46}\) Equation (3) represents an approach which is somewhat different from those discussed in Chapter II which disaggregate urban area output. Unfortunately, the available data do not permit much disaggregation (see Chapter IV). This failure to disaggregate increases the chances of rejecting increasing returns to scale in the urban area. Though the procedure adopted in this study is unaffected, it is a qualification that must be borne in mind when considering the results of Chapter VI.

\(^{47}\) The assumption of constant utilization rates between communities may be invalid. Presumably these rates will vary according to whether the community is growing or declining. This gives rise to the "regression fallacy" problem. (See Johnston, 1960, pp. 183-93.)

\(^{48}\) This development follows Berndt and Christensen (1973a), Burgess (1974), and Christensen, Jorgensen, and Lau (1973).
first is how to measure inputs and outputs, a subject which occupies most of Chapter IV. The second concerns the a priori specification of the functional form of the production function. The problem with using, say, a Cobb-Douglas production function is that it imposes, a priori, equal elasticities of substitution between inputs (which are equal to one) and homogeneity. In general, most production functions imply some maintained hypotheses and the objective here is to select a functional form which has a minimum number of prior restrictions. The function should be amenable to tests of the degree of return to scale and separability (equal elasticities of substitution). The Transcendental Logarithmic Production Function (the Translog) is one such function. The Translog can also be used to test for the Cobb-Douglas production function.

Assume there exists a technological relationship for urban value added which can be written:

\[(4) \quad \ln V = \ln A + G(\ln L, \ln K, \ln E).\]  

This implies that there exists a production function for V of the form \(V = A^*F(K, L, E)\). A second order approximation for the value of \(G\) in the neighbourhood about the point where all inputs are unity is given by taking a Taylor series expansion and omitting terms of a higher order:

\[(5) \quad \ln V - \ln A = G - G(0) + \sum_{i=1}^{n} \frac{\partial G}{\partial \ln x_i} \ln x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 G}{\partial \ln x_i \partial \ln x_j} \ln x_i \ln x_j ,\]  

\[\text{In time series analysis } A \text{ is an index of Hicks-neutral technological change. In a cross-section study } A = 1 \text{ (} \ln A = 0 \text{) if inputs and outputs are normalized at the median observation (using } V \text{ as the ranking variable) of a normal distribution. If the distribution is skewed then } A \neq 1.\]
where \( x_i \) represents the quantity of the \( i^{th} \) factor. In the three-input case,

\[
G = \ln \alpha_0 + \alpha_L \ln L + \alpha_E \ln E + \alpha_K \ln K
\]

\[
+ \frac{1}{2} Y_{LL} (\ln L)^2 + \frac{1}{2} Y_{EE} (\ln E)^2 + \frac{1}{2} Y_{KK} (\ln K)^2
\]

\[
+ \gamma_{LK} (\ln K \ln L) + \gamma_{LE} (\ln L \ln E) + \gamma_{EK} (\ln E \ln K)
\]

or,

\[
\ln V = \ln \alpha_0 + \alpha_L \ln L + \alpha_E \ln E + \alpha_K \ln K
\]

\[
+ \frac{1}{2} Y_{LL} (\ln L)^2 + \frac{1}{2} Y_{EE} (\ln E)^2 + \frac{1}{2} Y_{KK} (\ln K)^2
\]

\[
+ \gamma_{LK} (\ln K \ln L) + \gamma_{LE} (\ln L \ln E) + \gamma_{EK} (\ln E \ln K)
\]

Equation (7) is the equation used in this study since \( V \) is easier to measure than \( G \). In sum, \( V = AG \), individual inputs are transformed into the aggregate input \( G \) by a Translog input function and the aggregate input \( G \) is transformed into value added (\( V \)) by the scalar index \( A \).

Apart from symmetry, which has been imposed in writing equation (7), a well-behaved production function should exhibit positive marginal products and be quasi-concave at every data point. Positive marginal products for the Translog implies

\[
\frac{\partial \ln V}{\partial \ln x_i} = \frac{\partial \ln G}{\partial \ln x_i} = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln x_j > 0, \text{ for all } i.
\]

50\( MP_i = \frac{\partial V}{\partial x_i} = (\partial \ln V/\partial \ln x_i) * x_i/V \). Requiring \( MP_i \) positive implies \( \partial \ln V/\partial \ln x_i \) positive provided \( x_i \) is positive, which will generally be the case.
Quasi-concavity requires the bordered Hessian matrix of the production function be negative semi-definite. The first and second order partials of $V$ required for this computation are given by,

\[
F_r = \frac{G}{X_r} \left( \alpha_r + \sum_{j} \gamma_{rj} \ln X_j \right)
\]

\[
F_{rr} = \frac{G}{X_r^2} \left[ \gamma_{rr} + (\alpha_r + \sum_{j} \gamma_{rj} \ln X_j)^2 - 1 \right] (\alpha_r + \sum_{j} \gamma_{rj} \ln X_j)
\]

\[
F_{rs} = \frac{G}{X_r X_s} \left[ \gamma_{rs} + (\alpha_r + \sum_{j} \gamma_{rj} \ln X_j)(\alpha_s + \sum_{j} \gamma_{sj} \ln X_j) \right]
\]

Finally,

\[
\frac{\partial^2 \ln V}{\partial \ln X_i \partial \ln X_j} = \frac{\partial^2 \ln G}{\partial \ln X_i \partial \ln X_j} = \gamma_{ij} \geq 0.
\]

Since the Translog is a second order approximation and a quadratic, it is not globally well-behaved. For the moment, it will be assumed that the Translog provides a good representation of production possibilities for the observation set.

The parameters of the Translog can be estimated directly from equation (7) or with cost-share equations, or by using both together. The first approach is straightforward and is used in Chapter VI. The second creates some difficulty if non-constant returns prevail and necessitates a short digression on distribution theory. The second approach is used in Chapter VII.

\[51\text{See Burgess (1974, p. 112).}\]
The assumption of non-constant returns invalidates neo-classical distribution theory, i.e., paying factors their marginal social products does not exhaust output; at least one factor cannot receive its social marginal product. In the one factor case this problem can be resolved by paying the factor its average product. Unfortunately this solution cannot be extended to the n-factor case.

The distribution problem means that the marginal product of a factor \( MP_i \) is not, in general, equal to its factor price. Instead, let the \( i^{th} \) factor's marginal product equal its factor price plus a constant,

\[
(13) \quad MP_i = w_i + c_i \quad 52,53
\]

The cost share of the \( i^{th} \) input in the total input cost is

\[
(14) \quad M_i = w_i(X_i/G) = MP_i(X_i/G) - c_i(X_i/G).
\]

But \( \ln G/\ln X_i = (\partial G/\partial X_i)(X_i/G) \) and with (6), (8), and (12), (14) implies,

\[
(14') \quad M_i = \alpha_i + \sum_j^\text{n} \gamma_{ij} \ln X_j - c_i(X_i/G) \text{ for all } i.
\]

The cost shares always sum to one and, in the three-input case,

\[ M_1 + M_2 = 1 - M_3. \]

Equations (14') allow direct tests of hypotheses about factors receiving their marginal products. They also indicate the effects of urbanization on factor marginal products and factor payments. The value

\[ \textit{52} \text{ Assume all factor markets are perfectly competitive.} \]

\[ \textit{53} \text{ An alternative formulation is } MP_i = c_i w_i, \text{ but then the parameters cannot be identified.} \]
of $c_i$ indicates the distribution adjustment that takes place with non-
constant returns to scale; the values of $\alpha_i$ and $\gamma_{ij}$ indicate what happens
to factor marginal products as urbanization takes place.

There is a problem with this approach. For example, let the produc-
tion function be a three-input Cobb-Douglas, i.e., $\gamma_{ij} = 0$ for all $i$ and $j$.
This implies $(1 - \alpha_1 - \alpha_2 - \alpha_3) = -\sum_i c_i(X_i/G)$. The equations for $M_1$ and $M_2$
yield estimates of $\alpha_1$, $\alpha_2$, $c_1$, and $c_2$. This implies $\alpha_3 - c_3(X_3/G) \equiv k$, since
$\sum_i M_i = 1$. Either $\alpha_3$ or $c_3$ cannot be a free parameter. Since $\alpha_3$ can be
estimated from equation (7) $c_3$ must be the computed parameter which makes
the identity hold.

B. Hypothesis Tests

To test for a homothetic production function impose $n$ constraints,
where $n$ is the number of inputs, such that

$$
\begin{bmatrix}
\sum \gamma_{i1} \\
\vdots \\
\sum \gamma_{ni}
\end{bmatrix} = 
\begin{bmatrix}
0 \\
\vdots \\
0
\end{bmatrix}
$$

To test for the degree of returns to scale impose $\sum_i \gamma_{ij} = 0$ for all $i$ and

$\sum \alpha_i \leq 1$. As an intuitive rationale for this consider the Cobb-Douglas
production function $V = AK^{\alpha} L^{\beta} Y$. For the Cobb-Douglas $\gamma_{ij} = 0$ for all $i$ and $j$ and $\alpha + \beta + \gamma > 1$ implies increasing returns to scale. 54

54 More rigorously:
A final set of tests on separability between pairs of inputs can also be carried out. Necessary and sufficient conditions for inputs $i$ and $j$ to be globally separable from $k$ are that

$$\alpha_i \gamma_{jk} - \alpha_j \gamma_{ik} = 0,$$
$$\gamma_{im} \gamma_{jk} - \gamma_{jm} \gamma_{ik} = 0 \ , \ m = 1, \ldots, n.$$

When $\gamma_{jk}$ and $\gamma_{jm}$ are nonzero the separability conditions can be written as

$$\ln\lambda^t V = \ln \alpha_0 + \alpha_L \ln \lambda L + \alpha_K \ln \lambda K + \alpha_E \ln \lambda E$$
$$+ \frac{1}{2} \alpha_{LL} \ln (\lambda L)^2 + \frac{1}{2} \alpha_{KK} \ln (\lambda K)^2 + \frac{1}{2} \alpha_{EE} \ln (\lambda E)^2$$
$$+ \gamma_{KL} \ln \lambda L \ln \lambda K + \gamma_{KE} \ln \lambda E \ln \lambda K + \gamma_{LE} \ln \lambda L \ln \lambda E$$

$$t \ln \lambda = (\alpha_L + \alpha_K + \alpha_E) \ln \lambda + \frac{1}{2} (\gamma_{LL} + \gamma_{KK} + \gamma_{EE} + 2 \gamma_{LK} + 2 \gamma_{LE} + 2 \gamma_{KE})$$
$$+ (\gamma_{KL} + \gamma_{KE}) \ln \lambda \ln K + (\gamma_{LL} + \gamma_{LK} + \gamma_{LE}) \ln \lambda \ln L$$
$$+ (\gamma_{EE} + \gamma_{KE} + \gamma_{LE}) \ln \lambda \ln E.$$

A homothetic production function requires $\sum_i \gamma_{ij} = 0$ for all $i$. Thus, for a homothetic production function $t = \sum \alpha_i$. If the production function is non-homothetic then the returns to scale, $t$, is a function of $K, L, E$ if $\lambda$ is set equal to one. Only at the point where all inputs equal one is $t = \sum \alpha_i$. Thus, the degree of returns is a variable.
\[
\frac{\alpha_i}{\alpha_j} = \frac{\gamma_{ik}}{\gamma_{jk}} = \frac{\gamma_{im}}{\gamma_{jm}}, \ m = 1, ..., n. \ 55
\]

If inputs are separable then the elasticities of substitution between them are equal. For instance, if capital and labour are separable from land then \( \sigma_{KE} = \sigma_{LE} \) and the production function can be written as

\[
Z = F(V(K,L), E).
\]

This would imply that higher land taxes would not alter the optimum capital-labour mix in an urban centre. As well, measures designed to increase the labour force, rather than to increase the capital stock, will not have any differential effect on land values.

IV. DETERMINANTS OF THE HIERARCHY

Since the objective of estimating a production function is to determine the role of the physical factors of land, capital, and labour, it is important to control for different product mixes, different product costs, and aggregation phenomena. The hierarchy can be used for this purpose since it reflects these phenomena.

The three factors which generate the hierarchy and the resulting variance in community product mixes are: 56

1. different production costs
2. different transportation costs
3. different demand intensities.


56 Ignoring institutional and cultural differences. These factors determine the economic characteristics of a community which are used in Chapter V to construct the discrete hierarchy.
If transport costs and demand intensities are relatively equivalent for all products\(^{57}\), then production cost will be the sole determinant of the hierarchy. The crucial aspect is the volume of output at which the minimum average cost is achieved. The higher the volume, ceteris paribus, the larger the minimum market area. Firms which do not locate centrally so as to minimize transport costs and to maximize market access will be at a competitive disadvantage.

This analysis is an explanation of the market areas of different economic activities and how they affect city size. In the case where the minimum average cost volume of output is the crucial determinant of the minimum market area it will also be the main determinant of urban size.\(^{58}\) A large output requires large inputs of capital and labour and, therefore, larger cities. However, differences in transport costs and demand intensities will affect the size of market areas, the volume of output, and city size. Higher unit transport costs and inelastic demands promote smaller market areas. The introduction of reality complicates matters even more. For example, if the terrain of the area varies, affecting transportation costs, or if there are local variations in taste, affecting the distribution of demand but not necessarily of population, market areas will be affected. In its empirical reality the hierarchy reflects the interaction of all these forces.

---

\(^{57}\) Per unit transport costs are the same percentage of selling price, the number of units demanded per capita is the same, and the population is uniformly distributed.

\(^{58}\) Spatial competition is akin to monopolistic competition and production takes place at the point of tangency between the demand and average cost curves. This tangency does not coincide with the point of minimum average cost. The text ignores this complication.
The hierarchy argument can be summarized in Figure 6. The solid line represents the continuous hierarchy. The discrete hierarchy implies that there are groups of activities with the same, or similar, market areas. The communities with only small market area activities will be in the lowest hierarchy level. Communities in the highest hierarchy level have some activities which cover the entire range of market areas and which increase the average market area of these communities. The horizontal lines represent this discrete hierarchy.

Figure 6

![Graph showing average market area of activities in a city vs. city size]

In sum, the hierarchy reflects an ordering of market area size with consequent variations in the concentration of activities and urban area size. When the production function is estimated the hierarchy serves to control for product mix, cost differences (production and transport), and
demand variations. In the absence of a hierarchy variable the regression results could reflect any one of the above phenomena. Since the problem in this study is to determine the degree of returns to scale to physical factors, effects due to the hierarchy must be separated out. Though there may be other interpretations of what is here called a hierarchy variable, there is no question that some such procedure must be adopted. Ideally, all the relevant phenomena would be controlled on an observation by observation basis. But this is impossible and the argument is that the hierarchy reflects all these influences.

V. THE IMPLEMENTATION OF THE HIERARCHY

It has been argued that the urban hierarchy must be included in a model of urban areas; it remains to show how this may be done.

There are three approaches:

1. estimate equation (7) with the hierarchy included as an independent continuous variable,

2. estimate equation (7) with the hierarchy included as an independent discrete variable,

3. group the communities into hierarchical levels and estimate equations (7) and (14') for each level.

The first approach, with a continuous hierarchy, implies that each urban area produces the same good, i.e., there is only one production function and different sized urban centres are just like different sized factories producing the same good.\footnote{In this case an explicit continuous hierarchy variable serves as a size control variable, in addition to its other functions, and serves as a method for avoiding the regression fallacy. Another way to avoid this problem is to group the data but this is equivalent to the third approach. See J. Johnston (1960, pp. 183-93) for a discussion of the regression fallacy.} The second approach has an analogous interpretation...
but the effect of the hierarchy is to create a different production function for each hierarchy level, production functions which differ only in the constant. The last approach postulates different production functions for each hierarchy level where efficiency parameters and constants may vary. 60

The first two approaches treat the hierarchy as a shift parameter. The alternative is to include the hierarchy as an input in the production function, i.e., \( V = A*F(K,L,E,H) \). This procedure is equivalent to the inclusion of externalities in production functions and in consumer utility functions. 61 One interpretation of the parameter estimates for the terms involving the hierarchy is that they represent the effects of agglomeration economies. 62 With the hierarchy included as an input (or as a shift parameter), the estimates of the marginal products of capital, land, and labour are then their private marginal products. Since the hierarchy is an unpaid "factor" the payment of private marginal products may then exhaust output, i.e., if the summation of the private marginal products is equal to output. If such payments do not exhaust output, then the institutional arrangements of income distribution must make the necessary adjustments. It is thus possible to distinguish between increasing returns due to private economies of scale, i.e., summation of the private marginal products exceed output, and increasing returns because of the hierarchy (agglomeration

---

60 The construction of the continuous and discrete hierarchies is discussed in Chapters IV and V. The continuous hierarchy's ordering and interpretation is analogous to time-ordered data.

61 See, for example, Henderson and Quandt (1971, pp. 267-75).

62 This is, of course, an interpretation of the hierarchy variable no matter how it is included.
economies).

The prior expectation is that the production function exhibits increasing returns to scale when the hierarchy is excluded and at least constant returns to scale when it has been included. If the previous argument is correct then it is possible to distinguish the contributions of private sector economies and agglomeration economies to the total degree of returns to scale. This view of the hierarchy implies that the estimated equations are, in fact, an aggregated equation (1) where the hierarchy is a proxy variable for all the value added terms ($v_j$'s).\textsuperscript{63,64} This interpretation is easily extended to the discrete hierarchy.

Another approach is to group the communities and use dummy variables for each group or hierarchy level. This method estimates one equation for the entire sample and suggests that the efficiency parameters of the production function are identical throughout the hierarchy; only the constant, the technology index, changes.

Finally, the equations can be estimated for each hierarchy level by ordinary least squares or simultaneously. This approach is the most consistent with empirical evidence and theoretical argument. It assumes

\textsuperscript{63}It is possible to reverse the argument, i.e., the coefficients on the hierarchy terms indicate the effects of changing output mixes and the economies of scale associated with producing higher order goods. Then the continued existence of increasing returns to scale can be attributed to agglomeration economies. Unfortunately it does not, at the moment, appear possible to discriminate between these two views.

\textsuperscript{64}If the four-factor production function exhibits constant returns to scale then movements up the hierarchy are necessary for continuing urban growth. Self-sustaining growth is possible only if increasing returns exist for land, capital, and labour independently of the hierarchy, e.g., in the Cobb-Douglas case $a_L + a_K + a_E \geq 1$. Thus, this explanation is consistent with, and lends credence to, the antropomorphic historical descriptions of cities battling for regional and national economic primacy. For example, see Brown (1974, pp. 29-30, 24-25).
that a discrete hierarchy exists and suggests the possibility of different production function efficiency parameters for each hierarchy level. This is a reasonable hypothesis since the economic activity mix is changing as one moves up or down the hierarchy.\(^{65}\) Since the communities are part of the same urban system there will likely be non-zero correlations between the disturbance terms in each hierarchy level equations. Therefore it is appropriate to estimate a stacked system of hierarchy equations using Zellner's Seemingly Unrelated Regression Equations (SURE) estimator.\(^{66}\)

The interpretation of the SURE estimator is that if a community is producing, say, a larger output than one would expect, then nearby communities would produce less. The surplus output from the overproducing community would then be exported to the surrounding communities. Unfortunately, the off-diagonal elements of the variance-covariance matrix will vary with the order of the elements within each hierarchy level. With cross-section data there is no natural ordering of the data. One possibility is to order the communities so that they are matched with the nearest centre of the next highest hierarchy level. However, this solution is unique only in the case of two hierarchy levels. Another difficulty in constructing the covariance matrix is that there will be different numbers of communities in each group which means the smaller groups must be blown-up to the size of the larger by repeating some communities as indicated by the distance ordering. But this will change the relative number of occurrences of

\(^{65}\) The grouping method discussed and used in Chapter V minimizes the variety of activities in a group, i.e., the method groups by type and quantity of activities in a community.

\(^{66}\) J. Johnston (1972, ch. 7).
communities and will affect the diagonal elements of the covariance matrix. Because of these problems, SURE cannot be used for the direct estimation of the production function. This problem does not prevent the use of SURE in estimating the factor share equations.

The construction of the discrete hierarchy involves the grouping of communities. Two other reasons for grouping the data are, to reduce the spread of the observation points and, therefore, the chances of violating the monotonicity and convexity restrictions, and to test for the effects of urbanization on factor productivity and on income distribution. At the same time, it is also a test of the hypothesis of "one" or "more than one" urban area production functions.

VI. SUMMARY

This chapter has discussed the functional form of the model to be estimated in Chapter VI. The aggregation assumptions have been pointed out and the connection between this study's approach and previous literature has been indicated. The incorporation of the urban hierarchy into the model was discussed. This discussion yielded the following sets of equations to be estimated in Chapter VI and Chapter VII:

(A) The Continuous Model,

\[ \ln Z = G + H + e, \text{ where } H \text{ is a continuous hierarchy variable}; \]

(B) The Continuous-Discrete Model,

\[ \ln Z = G + H_1 + \ldots + H_n + e, \text{ where } H_1, \ldots, H_n \text{ are dummy variables for } n \text{ hierarchy levels}; \]
(C) The Discrete Model,

\[
\ln Z_1 = G_1 + e_1 \\
\vdots \\
\vdots \\
\vdots \\
\ln Z_n = G_n + e_1, \text{ for } n \text{ groups of communities with each equation estimated by OLS;}
\]

(D) Cost Share Equations,

(i) Equations (14') are estimated for the entire sample,

(ii) Equations (14') are estimated for each hierarchy level using the SURE estimator.

In all cases \( e \) is the disturbance term. All methods allow the imposition of the constraints discussed in this chapter.

Finally, grouping the urban centres is a method of (a) implementing the hierarchy, (b) testing for "one" or "more than one" production function, (c) controlling for the regression fallacy and other systematic variance, (d) making the aggregation assumptions more reasonable, (e) reducing the chances of violating the technical conditions which must be imposed on the translog production function, and (f) testing for the effects of urbanization on factor productivity and income distribution.
Chapter IV
THE DATA

I. INTRODUCTION

This chapter deals with the data used to construct the hierarchy and
to measure output, capital, land, and labour. Theoretical and practical
problems associated with various measures are discussed.

Section II considers the general problem of the appropriateness of
the data for testing the model of Chapter III. The third section outlines
the data used to construct the hierarchy. The measurement of output and
inputs is discussed in a fourth section which also considers the general
problem of defining the inputs and the output of an urban area. Most of
the discussion in section four is concerned with the problems of implementing
equation (7) of Chapter III, but part of this section is also devoted to
discussing the implementation of equations (14') in Chapter III. A summary
concludes the chapter.

II. THE APPROPRIATENESS OF THE DATA SET

This is a cross-section study which uses observations on 464
communities in Saskatchewan for the year 1971.67 The communities vary in
population from 30 to 139,469 and in aggregate non-farm income from
$17,767 to $362,647,420.

67The observation set consists of communities for which a complete
set of descriptive variables exists. It is defined as the intersection of
the incorporated communities listed in the Annual Report and in Dun and
Bradstreet and which can be related to the coded area of the Income Tax
tape and the 1971 census tape.
A natural question is whether this is an appropriate sample for the testing of urban models which postulate increasing returns to scale which result from agglomeration economies or from some other type of economy. For example, it is arguable that, in Saskatchewan, only Regina, Saskatoon, Moose Jaw, and Prince Albert are likely to generate significant agglomeration economies and, therefore, Saskatchewan data would not be adequate to test for agglomeration economies. Of course, as was indicated in Chapter III, the level of aggregation required in this study makes it difficult to distinguish increasing returns to scale due to economies internal to firms or industries from those due to agglomeration economies. But this study's significance does not rest on the determination of the relative importance of different factors in the creation of increasing returns to scale in an urban area. Rather, the crucial postulate is that, in both theory and practice, the economic activity of an urban area can be described by an aggregate urban area production function.

Most models of urban centres implicitly or explicitly postulate some type of production function. Whether one assumes constant or increasing returns to scale, or that the degree of returns to scale varies with urban size, or that the hierarchy is an important factor, the estimation of a production function for any urban area is appropriate. The estimated parameters may then be used to evaluate different assumptions.

III. DATA FOR CONSTRUCTING THE HIERARCHY

Both continuous and discrete hierarchies are used in this study. The usual method of generating a continuous hierarchy is to rank communities by population size. This approach is used in Chapter VI. Other univariate
possibilities would follow the same principle, a variable is selected and communities are ranked according to the value of the variable. Because of the high correlation of population with most economic variables the use of population alone is sufficient.

The construction of a discrete hierarchy requires the use of a clustering algorithm which groups the communities into hierarchy levels. A set of characteristics describing each observation serves as the input. It is important to have a relatively complete description of each community. The characteristics used in Chapter V of this study fall into two broad categories:

1) Private Sector Variables: The Dun and Bradstreet Reference Book, March 1971 lists community establishments by four-digit SIC codes. The Reference Book also lists the number of different banks operating in the community. Another source gives the number of grain elevators in each community. Also included are the number of doctors, lawyers, dentists, veterinarians, and engineers in each community.  

68 While relatively complete, the listing does omit some SIC activities and does understate the total number of outlets in particular categories.

69 This, as well as the number of bushels shipped and the value of grain shipments, is obtained from the Canadian Grain Commission's Summary of Primary Elevator Receipts at Individual Prairie Points.

70 These totals are obtained from the relevant professional societies and from the Provincial Department of Health.
(2) Public Sector Variables: This category covers functions operated, funded, or located by federal, provincial, or municipal authorities. It includes the number of (i) provincial government and crown corporation offices, (ii) liquor board stores or vendors, (iii) RCMP detachments and officers, (iv) hospitals and beds, (v) hospital related institutions and number of beds, and (vi) schools (primary, high school, composite, technical, university) located in each community. 71

In all, there are 406 different characteristics. However, the number of characteristics used in Chapter V is reduced to 378 by aggregating those variables which appear only in Regina into one variable and those which appear only in Saskatoon into another. Chapter V discusses the use of these 378 variables in the construction of the hierarchy.

IV. THE MEASUREMENT OF OUTPUT AND INPUTS

A. Output

The total value of factor payments or the value of final sales can be used to measure output. In this study the factor payments approach is used. While the crucial questions deal with what is to be included in the output and input measures, it is easier to answer them after considering some aspects of the output measure.

In a closed economy the total value of factor payments equals the

71 This information was obtained directly from the appropriate government departments.
value of output. If all factors filed income tax returns and were honest
then the income tax data could be used to measure output. Of course, the
assumptions of a closed economy and, perhaps, honesty are not fulfilled for
an urban area but the use of the income tax data appears to be the only way
of deriving a usable measure of urban area output. The data used in this
study are taken from the Saskatchewan Government's income tax tapes for 1971.72

The variable used to measure output is total net income (NY), i.e.,
total income (TY) minus foreign income (FY) minus dividends (DI).73 In this
context FY refers to income received from outside the country which cannot
be included in the output of local urban areas. The rationale for deducting
DI is not as clearcut. The assumption is that most dividends are received
from firms which operate outside of the province. This is reasonable for
all but the largest four or five communities. The bias will be small but
does increase the chances of rejecting the increasing returns to scale
hypothesis. The components of TY are employment income (wages and salaries),
net self-employment income, and net income from other sources such as rent
and investments. The "net" refers to income after operating costs. Thus,
TY approximates value added, which is the appropriate dependent variable
when estimating a production function.74 Farm income earned by residents

72 Confidentially restrictions prevented outputting income by all of
the nine categories (sources) to which income is attributed. Totals for
aggregates were obtainable and some of these are discussed below.

73 Since the correlation between TY and NY is .99 and FY is only $689,529
out of $1,344,084,413, it likely does not matter which is used. Even if
FY is concentrated in Regina and Saskatoon, each with over 300 million in
income, the effect would be small.

74 Assuming separability between "imported" materials and primary
factors.
of an urban area must be, and is, deleted.\textsuperscript{75} Of course, their influence on the size of urban output is reflected in their purchases of goods and services which generate income for urban suppliers. Also, TY does not have to be adjusted for transfer payments since these were not taxable in 1971 and are not included on the income tax tapes.

Since Saskatchewan and its communities do not represent a closed economy, significant inflows and outflows of income occur. However, after excluding farm and other resource income, one would expect a net outflow from the province.\textsuperscript{76} As well, most of the province's manufacturing, processing, and higher order service industries are located in the larger centres and there will be a net inflow of income from the smaller to the largest centres.\textsuperscript{77} This will result in an overstatement of returns to scale when one production function is estimated for the entire sample.

Since these biases exist for most of the variables it is appropriate to consider how to remove or reduce them. The most direct solution is to group the communities and to select an appropriate base to index the data.\textsuperscript{78}

\textsuperscript{75} This is easily done since taxfilers are identified by a category code, one of which is for farmers. This results in the deletion of all income earned by farmers, regardless of source.

\textsuperscript{76} This is simply because most manufactured products, and many services, are imported from outside the province.

\textsuperscript{77} And there is a net inflow to all urban centres from the country and a net inflow from the rest of the world to Saskatchewan's farm and resource industries.

\textsuperscript{78} J. Johnston (1972, pp. 192-194).
Since measurement biases and errors within community groups will tend to be similar the groups serve to control for these errors. As well, if output is affected by variables which cannot be controlled directly, or standardized within classes, then grouping gives better estimates of the parameters. For example, in the case of one uncontrolled variable (e.g., the quality of farm land surrounding a community) and two classes the difference in means ($\bar{y}_2 - \bar{y}_1$) overstates the true class difference (d) and a failure to adjust for classes results in an overstatement of the slope parameters ($a'' > a'$). This is illustrated in Figure 7. This procedure can also be used if there is any systematic relationship between the size of the urban area and the accuracy of the data obtained from the urban area.\footnote{A cursory examination of municipal budgets indicates that such a positive relationship between urban size and data accuracy may exist.}
B. Urban Area Output and Inputs: Definitions

Given that TY is the proposed measure of output it is necessary to consider how to interpret and to define the output and inputs of the urban area. In general an input in community \( i \) \( (x_i) \) is allocated between the private goods, housing, and public sectors:

\[
x_i = x_{p}^{i} + x_{h}^{i} + x_{t}^{i}.
\]

Clearly \( x_{p}^{i} \) must be included as an urban area input and the private sector's output must be included in the urban area's output. Private sector payments to labour, land, and capital owners, as well as profits, are all recorded when income tax returns are filed. The income measure of output therefore includes the private sector output.\(^80\)

The issue of \( x_{t}^{i} \) is more complex. The public sector presumably supplies basic infrastructure for the community, as such the public sector's output can be viewed as an intermediate product used in production by the private goods sector. Since the private sector pays the taxes which are used to provide the public services the public sector output will be included in TY which is evaluated before local taxes are paid. The problem is that some public sector factors, e.g., labour, file income tax returns and double-counting results. Thus, output is overstated and indexing the data solves the problem only if wages and salaries are the same proportion of municipal expenditures for all communities. An analysis of the budgets of approximately 25% of the 464 communities indicated that such an assumption

\(^80\)To the extent that non-resident owners receive payments the tax data will understate the urban area's output. Similarly, inflows will overstate the centre's output. Since nothing is known about the net result the assumption of zero net flows will be adopted.
does little violence to the facts.\textsuperscript{81}

Two other problems relate to the variance of the prices and quality of public goods across communities. Mera argues that price adjustments need not be made when comparing public services;\textsuperscript{82} however, the quality problem is more difficult. It is necessary to assume that quality differences are completely reflected in prices or that bundles of public services provided by different communities are not significantly different. The former seems more reasonable.

Similar problems exist for provincial and federal public goods since these governments provide services to the entire province but their inputs and factor payments are concentrated in a few centres. Thus, a regression for the entire sample could, because of this concentration, indicate increasing returns to scale.\textsuperscript{83} Again, grouping will help control for inter-community variation in these variables.

The housing sector is a more difficult problem. Urban models usually

\textsuperscript{81}The analysis compared the ratio of municipal expenditures on labour to total expenditures for a set of randomly selected communities. All the computed ratios varied between .1 and .4 with the higher values being associated with the larger communities. Appearances to the contrary, the observed range of .3 is not large since the larger communities generally supply a much more detailed analysis of expenditures with labour costs carefully noted. Smaller communities often cite only the total project cost without a detailed breakdown. Assuming that some labour is used and paid, and realizing that an extra hundred dollars in labour costs can double the ratio for a small community, it would seem that the difference in the ratios is not substantial. In any event, within groups (e.g., cities, towns, villages) the divergence disappears and grouping should eliminate the problem.

\textsuperscript{82}Mera (1973, pp. 311-12). This point is discussed below.

\textsuperscript{83}This would occur if the public goods produced by these governments exhibited increasing returns. Of course, the public sector inputs must be included in the input measures while the income payments should be deleted.
divide private aggregate production into a housing and a goods sector. The Saskatchewan data cannot be disaggregated to conform to this division. The obvious solution is to assume identical production functions for the two sectors and to estimate the appropriate function.

If housing is treated as an output then the value of housing services must be included in the output measure. If everyone rents housing or is paying off a mortgage then these payments would measure the value of housing services. These payments are recorded in the income tax data. Two difficulties are that (1) outright home ownership and (2) rental and mortgage payments to firms outside the community will result in an understatement of output. Because a separate assessment for the housing stock is unavailable an imputation of the value of housing services cannot be made. If the regression analysis indicates increasing returns for smaller communities then this is an "acceptable" bias. However, decreasing returns to scale could only be accepted with qualification. To some extent this problem is reversed for the larger urban areas. Therefore,

---

84 An exception is Dixit (1973).

85 Note that this differs substantially from Mills' assumption (Mills, 1967).

86 Assuming perfectly competitive markets and ignoring the consumption aspects attached to ownership.

87 This could be a serious problem. Conversations with provincial government researchers indicate that developers in Regina and Saskatoon often occupy monopoly positions in small communities.

88 An examination of the income data, which allows separate totals for rental income, indicates that the assumption of proportional rental flows is consistent with the data.
when the analysis is performed for the entire sample, returns to scale will be overstated.

The previous discussion suggests what may be a critical problem. Income flows into and out of communities bias the income measure of output if they are due solely to non-residency. If people work in larger centres but live in smaller centres then the observed income of large centres will be lowered and that of small centres raised. This would increase the chances of incorrectly rejecting the increasing returns to scale hypothesis. While a certain amount of inter-community income flows represents purchases of services from one community by another and are appropriately included in the supplying centre's output, it is impossible to disentangle the relative magnitudes of the two types of flows. For that matter, the total flow cannot be determined either. Again, grouping the communities serves as a partial control for the non-residency problem.

A similar problem arises with the cost-share equations. The natural measure of $M_i$, the cost share of the $i^{th}$ input in the total input cost, is the $i^{th}$ factor's share of total income. In the three factor case, given NY, data on two factors are required. But a taxfiler is categorized by his largest source of income, i.e., if he receives both investment and labour income he is placed in the category corresponding to his largest source. Thus, income shares ($M_i$) can be over- or understated. Some of the error will be offset by opposite categorizations but, a priori, prediction of the existence or direction of a net bias is difficult. $^89$ This is called the categorization problem.

$^89$ The assumption adopted for this study is that non-residency flows and categorization biases cancel.
A different issue is how wage and price differentials between communities affect the measurement of value added and of the cost share. However,

... in a country where a free market prevails, inter-regional or intercity price differentials contain meanings which should not be eliminated in a comparison. The price of an apple of a specified quality may be higher in a large city because the rent and wage the retailer has to pay are higher and also because people are willing to pay a higher price for it. At the same time, such a high-cost city is not isolated from the rest of the economy. It sells goods and services in exchange for imports. Therefore, a high price index is not an accident but a sustainable property of certain cities. A higher price for a certain good at a specific location implies a relatively higher value of the good at the specific location. The same argument holds for wages.90

In production function estimation the concern is with, say, the number of shirts produced, but value added is the usual measure of output. Different input and output prices imply different value added even for the same physical output. But, in a competitive interrelated market, there must be price-wage limits imposed by transportation costs; and the delivery of output to market must be part of producing the output for final consumption. The high-cost city producer charges a price commensurate with his delivery advantage and pays workers for their increased productivity and any other compensation required to attract labour. In sum, it is not clear that just physical output is relevant in the context of the city. What should be relevant is the final delivered product, and this output is measured appropriately by final selling price or total factor payments. Of course,

90Mera (1973, pp. 311-12).
the more imperfect the market is, the less valid the statement. But Saskatchewan is an integrated and open economy so that price differentials should not present a problem.

C. Measurement of the Inputs: Land, Capital, and Labour

(i) Land

The total land area is divided among the three sectors, i.e.,

\[ E^i = e_p^i + e_h^i + e_t^i \]

In the theoretical models the private goods producing sector is crucial. A strict implementation of this model requires measuring \( e_p^i \) which is also, in the theoretical model, the area of the CBD when all goods production is contiguous. However, the CBD's area cannot be measured and, in any case, is inappropriate when production is decentralized. The total incorporated land area of the community (\( E_l = E^i \)), as given in the Annual Report, must be used. Since the model employed here treats public and private sector inputs as aggregates, \( E_l \) is the correct measure of the land input.

The difference between empirical implementation and theoretical desiderata is minimized when the data are normalized. Consider an average community in a particular hierarchy level, i.e., a community whose income

---

91 Competitive markets insure that the value added measure is appropriate.

92 Vacant land is not considered since the estimation of a production function assumes a full utilization of factor service flows. In practice, the vacant land problem is reduced or eliminated when the data are normalized.

93 Mills (1967) and Dixit (1973).

94 In general, the variables \( L^i, E^i, \) and \( K^i \) are the ones which should be measured, and there is usually more than one way to measure each variable. (See below.)
is the median for that hierarchy level. If the area served by the entire hierarchy is homogeneous and if each urban centre has access to the same technology,\textsuperscript{95} then each centre in a particular hierarchy level will supply the same type of goods to its tributary area and land use patterns for communities within a hierarchy level will be similar.\textsuperscript{96} That is, $x^i_j/x^i$ would be the same for all communities in the same hierarchy level. Differences would be attributed to errors in optimization or different adjustment rates when relative prices change. In this case an index of total land area is equivalent to an index of $E_p$.

The problem is that $x^i_j/x^i$ may change with movements up or down the hierarchy. For the land input Dixit, in simulations of the optimum city,\textsuperscript{97} finds that $e^i_t/E^i$, the proportion of total land area in the public sector, declines as size increases. Since the CBD is a declining proportion of total land area, $e^i_h/E^i$, the proportion of total land area used in the housing sector, must be increasing. After the optimum size is reached the CBD

\textsuperscript{95}Suppose $n$ goods are produced in the hierarchy, then the production function is for a composite good consisting of these $n$ commodities, i.e.,

$$Z^i = \sum_{k=1}^{n} w_k z_k$$

where $w_k$ is the weight attached to good $z_k > 0$ for all $k$. Only centres in the highest hierarchy level produce all $n$ goods. Lower level centres devote resources to producing $m < n$ goods, i.e.,

$$Z^i = \sum_{k=1}^{m} w_k z_k$$

and $z_k = 0$ for $k = m+1, \ldots, n$. This is the economic difference between hierarchy levels and is the distinction which guides the construction of the hierarchy in Chapter V.

\textsuperscript{96}These assumptions seem reasonable for Saskatchewan. Implicit is the assumption that the communities were established at the same time (thereby facing the same relative prices); in fact, most communities in Saskatchewan were established between 1885-1910.

\textsuperscript{97}Dixit (1973, p. 646-49). Whether or not Dixit's simulations are relevant is a moot point. Empirical studies are inconclusive. See Bourne and Maher (1969) and Urban Land Use in Ontario.
remains a constant proportion of total land area. This suggests that using a mid-level hierarchy community as the base for the land index will overstate $e_p$ for higher hierarchy levels and understate it for lower levels.

Aside from the acreage measure of land there are two other possible land measures. They are the assessed value of land in the community, exclusive of exempt property (denoted as E2), and the assessed value of all land in the community (denoted as E3). Since E2 and E3 will be related to the value of the structures on land and since it is arguable that land prices (unlike capital and labour prices) will have a large variance over the Province, E1 is the preferred measure.

Non-resident ownership and transferred rental payments seem significant for land, even though the total rental and mortgage income is a small percentage (0.75%) of total income. The largest four cities have 17.8% of total land area, 32% of total land assessments, and 41.8% of rental payments. This could be a problem when estimating the cost share equations.

(ii) Capital

The capital used in the urban area should include both equipment and structures. Unfortunately a direct measure of equipment is not available. The Annual Report does list dollar value assessments for improvements (K1) and business (K2). An additional unpublished assessment for tax-exempt organizations is also available (K5). The assessment's base year is the same for all communities.

The variable K1 is for all private structures and is equal to 60% of the structure's gross value, where gross value is defined as the cost of replacing the structure today. The assessment K2 is an additional assessment on business establishments which is based on the number of square
feet occupied. \(^{98}\) If \(K_2\) is a proxy for equipment then \(K_3 = K_1 + K_2\) may be the appropriate capital measure. A final measure is \(K_4 = K_3 + K_5\) where \(K_5\) is the public sector counterpart of \(K_3\). Berndt and Christensen concluded that a consistent aggregate index exists for equipment and structures in United States manufacturing. \(^{99}\) \(K_3\) and \(K_4\) are two possible aggregates and are the preferred measures.

As in the case of land, non-resident owners of capital will affect the accuracy of the income measures of output. Investment income is 4.7% of total income and 68.5% of it is concentrated in the four largest cities. However, only 50% of total assessments for private structures, 47.5% of business assessments, and 49.5% of total income are concentrated in these same four cities.

(iii) Labour

The obvious measure of labour is simply the number of workers in an urban area, which is obtainable from the census tract data of the 1971 Canadian Census. This measure of labour (\(L_1\)) is unadjusted for quality (human capital) and the question is whether it should be.

The simplest solution to the quality problem is to argue that

---

\(^{98}\) The calculation of \(K_2\) is based on a survey of four-digit SIC establishments. The objective was to determine the profit of each establishment and then to allocate that profit to the different types of space used. For example, for millworks and lumber, office space was assessed at \$4.50/sq.ft., display at 2.25, equipment and work area at 1.50, work area and storage at 1.10, warehouse at .75, 2nd floor storage at .50, and yardage at .25. Since the estimated profit is net of variable costs and rent this assessment will be based on the returns to capital, especially equipment.

\(^{99}\) Berndt and Christensen (1973, pp. 99-100). Whether this result can be applied to Saskatchewan and whether simple addition is the appropriate aggregation procedure are moot points.
increased quantities of human capital in urban areas is an economy of agglomeration. Citing a study by Fuchs, Mera concludes "that the marginal productivity of labour increases as the size of the urban centre increases throughout the observed range." This can be explained by two hypotheses:

(1) the labour force in large cities is more productive because a large city attracts high-quality workers whose supply is independent of the size-distribution of cities,

(2) potential high-quality workers are given opportunities for greater development in large cities than in small because of the diverse opportunities and severe competition that are only available in large cities.

Hypothesis one requires a human capital or quality adjustment for any labour input measure; hypothesis two does not require a quality adjustment. Though Mera and Conrod opt for hypothesis two this study uses both adjusted and unadjusted data.

Human capital indices can be constructed by census tract and community from the 1971 Census. This involves weighting each educational category and computing a total which, when indexed by the median community, serves as multiplier to adjust the absolute quantities of labour. This

100 See Conrod (1973) and Mera (1973).


102 The educational categories are: grades 0-5, 5-8, 9-10, and 11-13, vocational courses, post-secondary non-university, two or fewer years of university, and 3 or more years of university. For grades 1-12 the midpoint of each class was used as the multiplier to estimate the total number of
quality adjusted labour variable is denoted as L2.

The best solution to the quality problem is related to the income distribution problem. Assume all factor markets are competitive, "... so that each activity pays the same price for a given factor. Furthermore, the wage rate, w, and the rental on capital, i, are assumed to be exogenous. These are the appropriate assumptions if the city's size is to be endogenous."\textsuperscript{103} Let capital and labour, for simplicity, receive their marginal products while land receives the residual. As human capital and productivity increase w will rise. If the relative amount of human capital increases as the urban centre's size increases, a higher w will be observed. Given a "raw" labour index and a wage index we can compute a quality adjusted labour index. The new index will be biased upwards to the extent that compensation payments for pollution, crime, cost of living, and congestion are made to attract labour to the urban area. This approach also requires that the Saskatchewan labour force be sensitive to real income differentials and be highly mobile.\textsuperscript{104}

To compute this wage adjusted labour measure (L3) the average wage per worker is computed for each community. When these averages are indexed they serve as the multipliers to adjust the raw labour measure. However such a measure is of questionable value since it is derived from the income

\textsuperscript{103} Mills (1967, pp. 202-203).

\textsuperscript{104} Courchene (1970) suggests that this is a realistic assumption.
data which also serves as the output measure. Accordingly L2 would seem superior to L3, but an a priori preference for L1 and L2 is nonexistent since each measure corresponds to a different hypothesis.

D. Measurement of $M_i$ and $G$

To estimate (14') of Chapter III requires measures of $M_i$ and $G$. 105 The value added or income measure can be used to measure $G$. $M_i$ must be measured by the factor income shares derived from the income tax data. However, as has been indicated, these measures are highly unreliable.

In any event, the cost share equations will be difficult to implement and interpret. At best, estimation of (14') will add to the confidence of the values obtained from estimating the production function directly; at worst the results from (14') will be rejected as unreliable.

V. SUMMARY

While data problems are severe they are not insurmountable. As indicated, grouping the data will ameliorate some of the problems. Even without the hierarchy it is clear that grouping is an empirical necessity.

The extent and number of biases in the data were discussed. The greatest difficulties are associated with the non-residency problem and income class categorization problem. Since most of these problems create a bias towards increasing returns to scale it is likely that any net bias is also in that direction.

105 $M_i$ is the cost share of the $i^{th}$ input in the total input cost while $G$ is the Translog aggregate input function.
Of the set of measures suggested and discussed the following are preferred:

(1) $K_3$: the sum of the assessments for all private structures plus the assessment on business establishments within a community.

$K_4$: $K_3$ plus the sum of the assessments for all tax-exempt structures within a community.

(2) $E_1$: the total incorporated land area of a community.

(3) $L_1$: the total number of workers in the community.

$L_2$: $L_1$ adjusted for human capital using labour force education data.

All four possible input sets are used in Chapter VI, but only $(L_1, E_1, K_3)$ is used in Chapter VII. ¹⁰⁶

¹⁰⁶ A sensitivity analysis was carried out on the other input measures mentioned in this Chapter. In general, the five variables listed above also seemed to be empirically superior to the others. This sensitivity analysis is not included in this study.
Chapter V

CONSTRUCTION OF THE DISCRETE HIERARCHY

I. INTRODUCTION

This chapter deals with the construction of a discrete hierarchy. Section II discusses the clustering algorithm and the criteria for selecting the optimal grouping. Part III contains the grouping which is finally selected and also includes some summary statistics describing the characteristics of each group.

II. THE CLUSTERING TECHNIQUE

This study uses the k-mean clustering technique which allows different weights to be attached to each of the characteristics used to describe a community.\textsuperscript{107} All the characteristics are standardized by their variance over the observation set. An observation is then placed in the group with the closest centre, where "closest" is determined by the smallest weighted and standardized n-dimensional Euclidean distance from a centre to the observation. In this study equal weights are attached to each dimension or characteristic.

The theoretical desiderata of the clustering algorithm are as follows:\textsuperscript{108}

(i) select $k$ initial random points as group centres,

(ii) add each new observation to the group whose centre is nearest,

\textsuperscript{107}MacQueen (1967).

\textsuperscript{108}MacQueen (1967) and Lance and Williams (1968).
(iii) adjust the centre of the group to which the observation is added,

(iv) after all the observations have been grouped, reclassify all observations using the means obtained from the first pass as the k centres.

The program can be elaborated to consider criteria under which k is increased or decreased as the observations are processed. The statistic which indicates whether k is optimal is the sum of the between-mean distances. The larger this is the greater the difference between the groups. However, results vary when the set of initial k centres is changed and there will generally be an overlap of the between-mean distances when the set of distances is compared for different numbers of initial centres, i.e., different k. The maximum of the minimums then determines the optimal k. If this procedure is determinant Figure 8 applies.

Figure 8

<table>
<thead>
<tr>
<th>sum of between mean distances</th>
</tr>
</thead>
</table>

k*     k (number of groups = number of initial centres)
Unfortunately, while a program which implemented the above procedure was available it could not be modified to handle the number of observations (464) and characteristics (378) involved. Two procedures are possible, either the number of characteristics could be aggregated to approximately sixty\textsuperscript{109} or a "stripped down" version of the program could be used. To minimize the information loss the latter procedure was chosen.

The rewritten procedure runs as follows:

(i) select $k$ centres \textit{a priori},

(ii) add each new point to the group with the closest centre,

(iii) (a) adjust the centre after a new observation is added to the group,
   (b) do not adjust the centre,\textsuperscript{110}

(iv) output the results and try another set of $k$ centres.

The rationale for the random selection of the initial centres and the moving of the mean is that the correct centres are not known \textit{a priori} and errors are corrected via the relocation procedure. The procedure used here has some counterbalancing advantages:

(1) other work and descriptive data, as well as theoretical arguments, allow reasonable \textit{a priori} selection of centres, and

(2) if the priors are correct then it makes little sense to adjust the centre as each new observation is added.

\textsuperscript{109} The aggregated variables are then used to group the communities. However, the matrix (464 x 378) was too large to be processed by available programs and preliminary subjective sorting would have been required. Accordingly, the variable aggregation procedure was rejected.

\textsuperscript{110} Of course, after all the observations have been grouped the centres must be adjusted before computing the between-mean distances.
Further, if the centres are selected appropriately then the results will be similar to those obtained from the more rigorous procedure. In the rewritten procedure, when option (iii)-(a) was used the results were extremely sensitive to the selection of the initial centres and to the order in which the points were added to the groups. Since option (iii)-(b) yields stable groups which are independent of the order in which the observations are considered, it is used to obtain the final groups.

III. THE RESULTS

To determine the optimal number of groups \((k^*)\) several clusterings, each using a different set of centres, were carried out for \(k\)'s from three to eight. The range of between-mean distances were:

<table>
<thead>
<tr>
<th>Between-Mean Distance</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k=3)</td>
<td>4.30</td>
<td>1.08</td>
</tr>
<tr>
<td>(k=4)</td>
<td>5.71</td>
<td>4.04</td>
</tr>
<tr>
<td>(k=5)</td>
<td>6.06</td>
<td>4.03</td>
</tr>
<tr>
<td>(k=6)</td>
<td>7.16</td>
<td>4.33</td>
</tr>
<tr>
<td>(k=7)</td>
<td>8.46</td>
<td>6.03</td>
</tr>
<tr>
<td>(k=8)</td>
<td>8.18</td>
<td>5.23</td>
</tr>
</tbody>
</table>

Using the maximum of the minimums criteria results in \(k^*=7\). This result is consistent with the findings of Stabler and Williams who obtained a six level hierarchy but whose sample set did not include the largest centres in the province. The selection of the best grouping with \(k=7\) was based on two criteria, a small total within-group sum of distances and a minimization of the number of anomalies within each group.\(^{111}\) These criteria resulted in

\(^{111}\) An anomaly is defined as a community which has either more than twice or less than half the number of different SIC types in the community used as the initial centre. Slightly more elastic bounds were used for Gl.
the groups of Table I. Approximately 17% of the communities can be considered 
anomalies, and two-thirds of these are located in the lowest hierarchy level. 
Some summary statistics of the seven groups are presented in Tables II 
through V.112

Table II contains the average values of output and inputs and their 
standard deviations, measured in terms of the median community (Harris), for 
each hierarchy level. As expected, output rises exponentially with the 
hierarchy. It is interesting to note the relative constancy of the standard 
deviations of the means. This suggests that weighted least squares need not 
be used.

Table II also indicates that L2 is consistently less than L1. On 
average it would appear that human capital is evenly distributed across the 
province, though the tendency is for relatively greater concentrations, on a 
per capita basis, in the lower hierarchy levels. This does not contradict 
the hypothesis that there are greater concentrations of human capital in 
larger centres since there are also greater concentrations of unskilled and 
uneducated labour. The net effect is lower per capita human capital in 
larger centres, i.e., the concentration of low-human-capital labourers 
relative to high-human-capital labourers is greater in larger centres. Thus, 
adjusting for human capital will not capture all of the agglomeration 
economies associated with urban areas; for the most part such economies 
are due to the greater diversity of labour, and inputs in general, rather 
than to quality changes. One implication of this for the regression 
analysis is that if L1 generates a degree of returns equal to x then L2 will 

112 Since there are only two observations in Group 7 they are included 
in Group 6. The variables are defined in Chapter IV: Y is output, L1 and 
L2 are labour measures, K3 and K4 are capital measures, and E1 is land.
<table>
<thead>
<tr>
<th>Group 6  (12)</th>
<th>Group 7  (23)</th>
<th>Group 8  (104)</th>
<th>Group 9  (92)</th>
<th>Group 10  (143)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regina1, Saskatoon</td>
<td>Holdfast1, Arcola, Beggough, Broadview, Bruce, Camrose, Chaplin, Cold Lake, Climax, Culp</td>
<td>Mendan1, Aberdeen, Anmore, Arrow Lake, Arvada, Austin, Avonlea, Beaverlodge, Bellrose</td>
<td>Sturgis1, Temiskaming, Thornhill, Tiny, Vandalia, Vancouver, Wakefield, Wapella, Webb</td>
<td>Camichael1, Ali1, Antler, Aylmer, Aylworth, Azquith, Baxton, Belle Plaine, Bethune</td>
</tr>
</tbody>
</table>
Table II  (At Harris all values equal one)
(Standard Deviations in Parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th>L1</th>
<th>E1</th>
<th>K3</th>
<th>L2</th>
<th>K4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire Sample</td>
<td>1.13</td>
<td>1.66</td>
<td>1.39</td>
<td>1.21</td>
<td>1.26</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>(3.71)</td>
<td>(3.41)</td>
<td>(2.42)</td>
<td>(3.22)</td>
<td>(3.25)</td>
<td>(3.27)</td>
</tr>
<tr>
<td>G7-6</td>
<td>66.97</td>
<td>81.13</td>
<td>24.40</td>
<td>56.20</td>
<td>51.61</td>
<td>49.65</td>
</tr>
<tr>
<td></td>
<td>(3.78)</td>
<td>(3.63)</td>
<td>(2.48)</td>
<td>(3.22)</td>
<td>(3.50)</td>
<td>(3.21)</td>
</tr>
<tr>
<td>G5</td>
<td>6.79</td>
<td>8.77</td>
<td>3.87</td>
<td>6.32</td>
<td>6.09</td>
<td>5.56</td>
</tr>
<tr>
<td></td>
<td>(2.41)</td>
<td>(2.27)</td>
<td>(2.41)</td>
<td>(2.14)</td>
<td>(2.20)</td>
<td>(2.13)</td>
</tr>
<tr>
<td>G4</td>
<td>2.08</td>
<td>3.00</td>
<td>1.68</td>
<td>2.19</td>
<td>2.18</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>(1.72)</td>
<td>(1.68)</td>
<td>(1.63)</td>
<td>(1.63)</td>
<td>(1.63)</td>
<td>(1.67)</td>
</tr>
<tr>
<td>G3</td>
<td>1.02</td>
<td>1.54</td>
<td>1.27</td>
<td>1.20</td>
<td>1.16</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>(2.25)</td>
<td>(2.14)</td>
<td>(1.99)</td>
<td>(1.92)</td>
<td>(2.03)</td>
<td>(1.96)</td>
</tr>
<tr>
<td>G2</td>
<td>.69</td>
<td>1.10</td>
<td>1.00</td>
<td>.77</td>
<td>.80</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td>(2.39)</td>
<td>(2.25)</td>
<td>(1.72)</td>
<td>(2.03)</td>
<td>(2.14)</td>
<td>(2.04)</td>
</tr>
<tr>
<td>G1</td>
<td>.52</td>
<td>.78</td>
<td>.97</td>
<td>.55</td>
<td>.63</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>(2.34)</td>
<td>(2.16)</td>
<td>(1.70)</td>
<td>(1.84)</td>
<td>(2.14)</td>
<td>(1.92)</td>
</tr>
</tbody>
</table>

generate a degree of returns equal to \( x + y, y \geq 0 \). The relationship between K3 and K4 is similar to that of L1 and L2. Apparently, public capital is distributed more evenly than private capital. This is not surprising since governments are committed to supplying a complete set of public services to all their citizens while the private sector is not.

Table III shows the rate of change of outputs and inputs from hierarchy level to hierarchy level and the average output of each input for each hierarchy level. The steady rise of the average productivity of land (Y/E1) and capital-adjusted labour (Y/L2) is explained by their slower rates of growth relative to capital (K3) and unadjusted labour (L1). An examination of Table IV indicates what is happening. Both labour-land ratios and the capital-land ratio rise dramatically while L1/K3 remains roughly constant and L2/K3 declines steadily. The constancy of K3/K4 means
Table III

A. Rate of Change From Level to Level

<table>
<thead>
<tr>
<th></th>
<th>Y</th>
<th>L1</th>
<th>E1</th>
<th>K3</th>
<th>L2</th>
<th>K4</th>
</tr>
</thead>
<tbody>
<tr>
<td>G7-6/G5</td>
<td>9.86</td>
<td>9.25</td>
<td>6.30</td>
<td>8.91</td>
<td>8.47</td>
<td>8.39</td>
</tr>
<tr>
<td>G5/G4</td>
<td>3.26</td>
<td>2.92</td>
<td>2.30</td>
<td>2.89</td>
<td>2.79</td>
<td>2.91</td>
</tr>
<tr>
<td>G4/G3</td>
<td>2.04</td>
<td>1.95</td>
<td>1.32</td>
<td>1.83</td>
<td>1.88</td>
<td>1.80</td>
</tr>
<tr>
<td>G3/G2</td>
<td>1.47</td>
<td>1.40</td>
<td>1.27</td>
<td>1.55</td>
<td>1.45</td>
<td>1.58</td>
</tr>
<tr>
<td>G2/G1</td>
<td>1.34</td>
<td>1.41</td>
<td>1.03</td>
<td>1.40</td>
<td>1.27</td>
<td>1.40</td>
</tr>
</tbody>
</table>

B. Average Output

<table>
<thead>
<tr>
<th></th>
<th>Y/L1</th>
<th>Y/E1</th>
<th>Y/K3</th>
<th>Y/L2</th>
<th>Y/K4</th>
</tr>
</thead>
<tbody>
<tr>
<td>G7-6</td>
<td>.825</td>
<td>2.74</td>
<td>1.19</td>
<td>1.30</td>
<td>1.35</td>
</tr>
<tr>
<td>G5</td>
<td>.774</td>
<td>1.75</td>
<td>1.07</td>
<td>1.11</td>
<td>1.22</td>
</tr>
<tr>
<td>G4</td>
<td>.693</td>
<td>1.24</td>
<td>.95</td>
<td>.95</td>
<td>1.09</td>
</tr>
<tr>
<td>G3</td>
<td>.660</td>
<td>.80</td>
<td>.85</td>
<td>.88</td>
<td>.96</td>
</tr>
<tr>
<td>G2</td>
<td>.631</td>
<td>.69</td>
<td>.90</td>
<td>.86</td>
<td>1.03</td>
</tr>
<tr>
<td>G1</td>
<td>.667</td>
<td>.53</td>
<td>.94</td>
<td>.83</td>
<td>1.08</td>
</tr>
<tr>
<td>Entire Sample</td>
<td>.681</td>
<td>.813</td>
<td>.934</td>
<td>.897</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Table IV

Input Ratios

<table>
<thead>
<tr>
<th></th>
<th>L1/E1</th>
<th>L1/K3</th>
<th>L1/L2</th>
<th>L2/K3</th>
<th>L2/E1</th>
<th>K3/E1</th>
<th>K3/K4</th>
</tr>
</thead>
<tbody>
<tr>
<td>G7-6</td>
<td>3.33</td>
<td>1.44</td>
<td>1.57</td>
<td>.82</td>
<td>2.12</td>
<td>2.30</td>
<td>1.13</td>
</tr>
<tr>
<td>G5</td>
<td>2.27</td>
<td>1.39</td>
<td>1.44</td>
<td>.96</td>
<td>1.57</td>
<td>1.63</td>
<td>1.14</td>
</tr>
<tr>
<td>G4</td>
<td>1.79</td>
<td>1.37</td>
<td>1.38</td>
<td>.995</td>
<td>1.30</td>
<td>1.30</td>
<td>1.15</td>
</tr>
<tr>
<td>G3</td>
<td>1.21</td>
<td>1.28</td>
<td>1.33</td>
<td>.97</td>
<td>.91</td>
<td>.94</td>
<td>1.13</td>
</tr>
<tr>
<td>G1</td>
<td>.80</td>
<td>1.41</td>
<td>1.24</td>
<td>1.15</td>
<td>.65</td>
<td>.57</td>
<td>1.15</td>
</tr>
<tr>
<td>Entire Sample</td>
<td>1.19</td>
<td>1.37</td>
<td>1.32</td>
<td>1.04</td>
<td>.906</td>
<td>.871</td>
<td>1.14</td>
</tr>
</tbody>
</table>
that (1) all the input ratios with respect to K4 follow the same pattern as those with respect to K3, except that they are 14% higher and (2) the regression results should be insensitive to the use of either K3 or K4.

The last Table contains the ordinary least squares estimates of $\alpha_i$ for production function $Y = A X_i^\alpha$; $i = L1$, $L2$, $E1$, $K3$, $K4$. The estimates are for the entire sample and each hierarchy level. Both labour variables show a rising productivity with movements up the hierarchy, though $L2$'s progress is more erratic. The estimates of $\alpha_k$ move up and down but are not significantly different from each other. Except for land, the $\alpha_i$ estimates are highest in G5 which suggests that this is the growth hierarchy level. G7-6 also appears to have growth potential while the situation in the other four hierarchy levels does not appear as promising. However, such conjectures should be held in abeyance until the more thorough regressions of Chapter VI are discussed.

<table>
<thead>
<tr>
<th>Table V</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Standard Errors in Parentheses)</td>
</tr>
<tr>
<td>L1</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Entire Sample</td>
</tr>
<tr>
<td>G7-6</td>
</tr>
<tr>
<td>G5</td>
</tr>
<tr>
<td>G4</td>
</tr>
<tr>
<td>G3</td>
</tr>
<tr>
<td>G2</td>
</tr>
<tr>
<td>G1</td>
</tr>
</tbody>
</table>
IV. CONCLUSION

In this chapter the empirical derivation of a discrete urban hierarchy was discussed. Though technical conditions forced the use of a second-best k-mean algorithm, it was argued that the results obtained would approximate the first-best solution. The seven-level hierarchy which was obtained is generally consistent with prior expectations and there are remarkably few anomalies. The seven groups obtained are reduced to the six in Tables II to V for use in the regression analysis of Chapter VI.
Chapter VI

ESTIMATION OF THE MODEL

I. INTRODUCTION

This chapter deals with the estimation of the Translog production function. The estimated equations are tested for the special case of a Cobb-Douglas technology, for homogeneity and the degree of returns to scale, and for linear and non-linear separability. The input measures used are:

1. \( L1 \) : the total labour force in a community,
   \( L2 \) : \( L1 \) adjusted for human capital,

2. \( E1 \) : the land area of a community,

3. \( K3 \) : private sector capital, and
   \( K4 \) : private sector and public sector capital.

Thus, there are four sets of input variables, \((L1,E1,K3)\), \((L1,E1,K4)\), \((L2,E1,K3)\), and \((L2,E1,K4)\), to be tested.

The production functions are estimated for each different method of implementing the hierarchy. The continuous and discrete hierarchy variables, models A and C of Chapter III, are statistically significant, i.e., they result in a significant decrease in the sum of squared residuals when compared to production functions without the hierarchy, while the continuous-discrete hierarchy, model B of Chapter III, is insignificant. To some extent, the degree of returns to scale depends on the input set used but the accepted homogeneity and separability hypotheses are the same for each variable set. Section II-A outlines the testing procedure; Section II-B discusses the continuous hierarchy; Section II-C briefly considers the continuous-discrete hierarchy; and, Section II-D focuses on the discrete hierarchy. Since \( K4 \) generates results which are highly similar to those
for K3, the variable sets with K4 are used only in Section II-B. The final part of Section II discusses the implications of the results. Section III summarizes and concludes the chapter.

II. ESTIMATION OF THE PRODUCTION FUNCTION

A. Introduction

This section discusses the results of the production function estimation. Part B considers the continuous hierarchy estimates, Part C deals with the continuous-discrete hierarchy, and Part D deals with the discrete hierarchy. These three parts are restricted to an analysis of the statistical tests; Part E summarizes these results and discusses their implications.

The test procedure for each set of regressions for \((L1,E1,K3)\), \((L1,E1,K4)\), \((L2,E1,K3)\), and \((L2,E1,K4)\) is as follows:

(a) a comparison of the Cobb-Douglas (CD) and Translog production functions is made,

(b) tests of the returns to scale and homogeneity hypotheses are made;

(1) tests for homogeneity, i.e., a constant degree of returns to scale,

\[
\begin{bmatrix}
3 \\
\sum_{i=1}^{3} y_{i1}
\end{bmatrix} \times \begin{bmatrix}
0 \\
\sum_{i=1}^{3} y_{i2}
\end{bmatrix} = \begin{bmatrix}
0 \\
\sum_{i=1}^{3} y_{i3}
\end{bmatrix}
\]
(2) tests for linear homogeneity, i.e., constant
returns and homogeneity,

\[
\begin{bmatrix}
\sum_{i=1}^{3} \alpha_i \\
\sum_{i=1}^{3} \gamma_{i1} \\
\sum_{i=1}^{3} \gamma_{i2} \\
\sum_{i=1}^{3} \gamma_{i3}
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix}
\]

(c) the separability hypotheses are tested.

In constructing the data indices the output measure is used first to
order the data, then to locate the median community, in this case, Harris.
All input and output measures are then normalized in terms of Harris, i.e.,
at Harris all \( X_i = 1.0 \). This procedure is also followed for the continuous
hierarchy variable. Since the continuous hierarchy deals with ranks
(population serving as the ranking variable) the value of the hierarchy
variable for a community is set equal to that community's rank. For
example, the community with the smallest population has a rank equal to one
while the largest has a rank equal to \( n \), the number of communities. In
this study the hierarchy variable ranges from \( \log(1/233) \) to \( \log(464/233) \),
where 233 is Harris' rank. Finally, the observations are ordered by

\[\text{An alternative is to construct the hierarchy variable using population instead of rank. This preserves order while allowing a greater variance in values--from } \log(22/254) \text{ to } \log(139469/254), \text{ where Harris' population is} \]
income and the Durbin-Watson statistic is interpreted as a test for omitted
variables (specification errors).

B. The Continuous Hierarchy

(i) Unadjusted Labour (L1)

The continuous hierarchy is represented by the rank variable (H1).

In Table VI, columns (1) and (2) contain the estimates of the Translog
production function for the variables sets (L1,E1,K3) and (L1,E1,K4). Both
equations use the hierarchy as a shift variable, i.e., the estimated
production function is \( V = A^*H^*H*F(K,L,E) \). In both cases the hierarchy
variable is significant at the 1% level, with \( F(1,453) = 24.69 \) and
\( F(1,453) = 21.15 \). The Cobb-Douglas equation is rejected at the 1% level for
both equations (with \( F(6,453)'s \) of 4.67 and 4.91 respectively). Linear
homogeneity is also rejected at the 1% level for both equations (\( F(4,453)'s \)
of 4.31 and 3.95 respectively). Finally, the hypothesis of homogeneity (a
constant degree of returns to scale) is rejected at the 1% level for both
equations with computed \( F(3,453)'s \) of 5.29 and 4.05.

254. This population variable produces a higher \( R^2 \), a higher t-statistic,
and lowers the estimates of the \( \alpha_i 's \) when compared to the rank variable.
(See Appendix II.) These results are not surprising when it is noted that
the correlation between the value added measure and the rank hierarchy
variable is .7469 while that for value added and the population hierarchy
variable is .9629. Accordingly, the rank variable is used since it seems
unwise to bias the results towards accepting the hypothesis that increasing
returns to scale (should they exist) are due to the hierarchy.

\(^{114}\) For future reference, the 1% and 5% significance levels which are
relevant in this study are:

<table>
<thead>
<tr>
<th></th>
<th>1%</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>F(1, ∞)</td>
<td>6.63</td>
<td>3.84</td>
</tr>
<tr>
<td>F(2, ∞)</td>
<td>4.61</td>
<td>3.00</td>
</tr>
<tr>
<td>F(3, ∞)</td>
<td>3.78</td>
<td>2.60</td>
</tr>
<tr>
<td>F(4, ∞)</td>
<td>3.32</td>
<td>2.37</td>
</tr>
<tr>
<td>F(5, ∞)</td>
<td>3.02</td>
<td>2.21</td>
</tr>
<tr>
<td>F(6, ∞)</td>
<td>2.80</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>1. (L1,E1,K3)</td>
<td>2. (L1,E1,K4)</td>
</tr>
<tr>
<td>-----</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>αL</td>
<td>.362035</td>
<td>.269503</td>
</tr>
<tr>
<td></td>
<td>(.0772401)</td>
<td>(.105959)</td>
</tr>
<tr>
<td>αE</td>
<td>.0314047</td>
<td>.0298670</td>
</tr>
<tr>
<td></td>
<td>(.0417509)</td>
<td>(.0512616)</td>
</tr>
<tr>
<td>αK</td>
<td>.500618</td>
<td>.570959</td>
</tr>
<tr>
<td></td>
<td>(.0898517)</td>
<td>(.120586)</td>
</tr>
<tr>
<td>γLL</td>
<td>.443139</td>
<td>.513732</td>
</tr>
<tr>
<td></td>
<td>(.173314)</td>
<td>(.194695)</td>
</tr>
<tr>
<td>γEE</td>
<td>-.0222547</td>
<td>-.0415339</td>
</tr>
<tr>
<td></td>
<td>(.0678129)</td>
<td>(.0677914)</td>
</tr>
<tr>
<td>γKK</td>
<td>.197215</td>
<td>.239226</td>
</tr>
<tr>
<td></td>
<td>(.208038)</td>
<td>(.221806)</td>
</tr>
<tr>
<td>γKE</td>
<td>-.0108752</td>
<td>-.0275766</td>
</tr>
<tr>
<td></td>
<td>(.0891054)</td>
<td>(.0924751)</td>
</tr>
<tr>
<td>γKL</td>
<td>-.297553</td>
<td>-.360618</td>
</tr>
<tr>
<td></td>
<td>(.175835)</td>
<td>(.195102)</td>
</tr>
<tr>
<td>γLE</td>
<td>-.0139700</td>
<td>-.0292165</td>
</tr>
<tr>
<td></td>
<td>(.0897145)</td>
<td>(.0929844)</td>
</tr>
<tr>
<td>C</td>
<td>.215607</td>
<td>-.118664</td>
</tr>
<tr>
<td></td>
<td>(.2091147)</td>
<td>(.0369389)</td>
</tr>
<tr>
<td>aH</td>
<td>.169986</td>
<td>.158064</td>
</tr>
<tr>
<td></td>
<td>(.0342136)</td>
<td>(.034734)</td>
</tr>
<tr>
<td>γHH</td>
<td>.0937790</td>
<td>.0937790</td>
</tr>
<tr>
<td>γHK</td>
<td>.069283</td>
<td>.069283</td>
</tr>
<tr>
<td>γHL</td>
<td>.128923</td>
<td>.128923</td>
</tr>
<tr>
<td>γHE</td>
<td>.124326</td>
<td>.124326</td>
</tr>
<tr>
<td>R²</td>
<td>.9395</td>
<td>.9395</td>
</tr>
<tr>
<td>SSR</td>
<td>48.0785</td>
<td>47.8388</td>
</tr>
<tr>
<td>DW</td>
<td>1.94</td>
<td>1.98</td>
</tr>
<tr>
<td>N</td>
<td>464</td>
<td>464</td>
</tr>
</tbody>
</table>

TABLE VI - *Non-Linear Constraint. Standard Error not Computed - Standard Errors in Brackets

* aL = aKE⁻¹, aE = aKE⁻¹, aK = aKE⁻¹, aL = aKE⁻¹, aE = aKE⁻¹, aK = aKE⁻¹
An alternative specification is to use the continuous hierarchy variable as an input in the production function, i.e., \( V = A^*F(L,E,K,H) \). The Translog estimates for this equation, using \((L_1,E_1,K_3)\), are presented in column (3) of Table VI. There is a significant reduction in the sum of squared residuals (SSR) when compared with the Translog in column (1) (the computed \( F(4,449) \) is 5.61). Both linear homogeneity and homogeneity, with respect to all four inputs, are rejected at the 1% level (\( F(5,449) = 8.39 \) and \( F(4,449) = 5.16 \)). Finally, linear homogeneity and homogeneity with respect to just the three physical inputs, are also rejected at the 1% level (with \( F(4,449) = 5.76 \) and \( F(3,449) = 6.45 \)).

If the hierarchy is excluded from the equation, neither the Cobb-Douglas nor homogeneity can be rejected (\( F(6,454) = 1.91 \) and \( F(3,454) = 1.68 \)). However, as is noted above, when the hierarchy is included homogeneity is rejected. It is apparent that the hierarchy, i.e., a community's position in the hierarchy, alters the production function in a non-neutral manner. As is clear from Exhibits A and C, the hierarchy (if it can be interpreted in a manner analogous to technical change) is capital-saving or labour-using in both the Hicks and Harrod sense. This could reflect either a changing output mix or different prices for inputs and output between communities. The acceptance of homogeneity for all six levels of the Discrete Hierarchy

---

115 See Chapter III, section V.

116 Hahn and Matthews (pp. 48-50, 1969). In the Hicks sense, if \( K/L \) is constant then capital-saving technical progress raises the marginal product of labour relative to capital; for Harrod, technical progress is capital-saving if the output-capital ratio is constant and the marginal product of capital falls (or, if \( MP_k \) is constant and the output-capital ratio rises). An examination of Exhibit A of this chapter and of Table IV indicates that the results are consistent with a capital-saving hierarchy.
lends further support to such suggestions.\textsuperscript{117}

The rejection of homogeneity means that the measure of returns to scale is a function of the input levels.\textsuperscript{118} As is illustrated in Exhibit A, within the range of observed input levels, the degree of returns increases with urban size. At the median centre, for the equations in columns (1) and (3), decreasing returns to physical inputs prevail ($t = 2.39$ and $t = 3.71$) but increasing returns to physical inputs and the hierarchy is accepted ($t = -1.94$ and $t = -3.96$), i.e., $\alpha_L + \alpha_E + \alpha_K + \alpha_H > 1$.

Exhibit A contains the logarithmic marginal products of labour, land, and capital computed at the mean values of the six hierarchy levels obtained in Chapter V. The parameter estimates from columns (1) and (3) of Table VI are used.\textsuperscript{119} The logarithmic marginal product of the $i^{th}$ input ($X_i$) is

$$M_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln X_j; \quad n=3 \text{ for column (1) and } n=4 \text{ for column (3).}$$

If the sum of the logarithmic marginal products equals one, then paying factors their marginal products will exhaust output. If the total exceeds one then increasing returns prevail and at least one factor must receive less than its marginal product; if the total is less than one decreasing returns prevail and at least one factor must receive more that its marginal product.

\textsuperscript{117}See Section II-D below.

\textsuperscript{118}When the Cobb-Douglas is used homogeneity is assumed a priori and the estimated equation can be tested for a constant degree of returns to scale. When the hierarchy is excluded from the equation increasing returns to physical inputs cannot be rejected; when the hierarchy is included constant returns to physical inputs is accepted ($t = -1.60$). However, increasing returns to physical inputs and the hierarchy is accepted ($t = -5.22$).

\textsuperscript{119}The pattern of logarithmic marginal products is similar for all the input combinations. Accordingly, the text presents and discusses the computations for just ($L1, E1, L3$) in Exhibit A and for ($L2, E1, K3$) in Exhibit C.
Exhibit A

Logarithmic Marginal Products*
(L1,E1,K3)

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>E</th>
<th>K</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>G76</td>
<td>1.152</td>
<td>-0.022</td>
<td>0.033</td>
<td>1.164</td>
</tr>
<tr>
<td>G5</td>
<td>0.795</td>
<td>0.012</td>
<td>0.236</td>
<td>1.043</td>
</tr>
<tr>
<td>G4</td>
<td>0.623</td>
<td>0.027</td>
<td>0.335</td>
<td>0.985</td>
</tr>
<tr>
<td>G3</td>
<td>0.503</td>
<td>0.030</td>
<td>0.411</td>
<td>0.944</td>
</tr>
<tr>
<td>G2</td>
<td>0.482</td>
<td>0.036</td>
<td>0.421</td>
<td>0.938</td>
</tr>
<tr>
<td>G1</td>
<td>0.430</td>
<td>0.035</td>
<td>0.456</td>
<td>0.921</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>E</th>
<th>K</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>G76</td>
<td>1.271</td>
<td>-0.172</td>
<td>-0.0723</td>
<td>1.03</td>
</tr>
<tr>
<td>G5</td>
<td>0.742</td>
<td>0.068</td>
<td>0.208</td>
<td>1.018</td>
</tr>
<tr>
<td>G4</td>
<td>0.520</td>
<td>0.084</td>
<td>0.347</td>
<td>0.947</td>
</tr>
<tr>
<td>G3</td>
<td>0.355</td>
<td>0.100</td>
<td>0.395</td>
<td>0.849</td>
</tr>
<tr>
<td>G2</td>
<td>0.414</td>
<td>0.001</td>
<td>0.358</td>
<td>0.773</td>
</tr>
<tr>
<td>G1</td>
<td>0.391</td>
<td>-0.015</td>
<td>0.356</td>
<td>0.731</td>
</tr>
</tbody>
</table>

*Evaluated at mean values of the groups, see Table II.

For both equations the logarithmic marginal products of capital and land move together and decline as the higher hierarchy levels are reached. The logarithmic marginal product of labour moves in the opposite fashion and is a rising proportion of the total. The two equations differ in the bottom three hierarchy levels where, for column (3), labour is initially the more significant factor, becomes less so, then "bottoms out" in G3. The lower totals for column (3) reflect the greater role of the urban hierarchy for that specification of the production function. As Exhibit A illustrates, the degree of returns increases with movements up the hierarchy or with increases in urban size.

Finally, the marginal product of a factor is \( MP_i = \frac{V}{X_i}(\alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln X_j) \). If an input's payment is systematically related to its marginal product then either its relative share in urban output changes (since its marginal product changes with urban size) or \( V/X_i \) must vary. The data of Table III,
and the computations in Exhibit A support the former hypothesis. However, if factor rewards are to be equalized between urban areas (as they must be when the system is in long-run equilibrium), then it is likely that, in the upper levels of the hierarchy, labour receives less than its marginal product. These hypotheses are tested in Chapter VII.

With constant returns and homogeneity rejected, the separability conditions are easier to impose. The relevant parameter restrictions are derived in Appendix I, Part B. The results of testing columns (1) and (2) for the three types of separability are summarized in Exhibit B.

None of the separability conditions can be rejected for either variable set at either the 1% or 5% level. However, for both variable sets, the equations which impose $\sigma_{LK} = \sigma_{KE} = 1, \sigma_{EL} = \sigma_{KL} = 1$, and $\sigma_{EL} = \sigma_{KL} \neq 1$ generate unacceptable results when tested for monotonicity and convexity.121

<table>
<thead>
<tr>
<th></th>
<th>column (1)</th>
<th></th>
<th>column (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MP_L</td>
<td>MP_E</td>
<td>MP_K</td>
</tr>
<tr>
<td>G76</td>
<td>.9604</td>
<td>-.0603</td>
<td>.0392</td>
</tr>
<tr>
<td>G5</td>
<td>.6153</td>
<td>.0210</td>
<td>.2525</td>
</tr>
<tr>
<td>G4</td>
<td>.4320</td>
<td>.0335</td>
<td>.3183</td>
</tr>
<tr>
<td>G3</td>
<td>.3320</td>
<td>.0240</td>
<td>.3490</td>
</tr>
<tr>
<td>G2</td>
<td>.3040</td>
<td>.0248</td>
<td>.3789</td>
</tr>
<tr>
<td>G1</td>
<td>.2870</td>
<td>.0186</td>
<td>.4286</td>
</tr>
</tbody>
</table>

120 The computed marginal products for column (1) and column (3):

121 Using the formulas in Chapter III, the marginal product of each input and the bordered Hessian of the production function are evaluated for each community (data point).
Exhibit B

(The F-scores in brackets are for the input set L1, E1, and K4; the other F-scores are for L1, E1, and K3)

<table>
<thead>
<tr>
<th></th>
<th>Linear ($\sigma_{ij} = 1$)</th>
<th>Non-linear ($\sigma_{ij} \neq 1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LK-E(\sigma_{LE} = \sigma_{KE})$</td>
<td>0.012 (.121)</td>
<td>.020 (.097)</td>
</tr>
<tr>
<td>$LE-K(\sigma_{LK} = \sigma_{KE})$</td>
<td>1.43 (1.73)</td>
<td>.058 (1.69)</td>
</tr>
<tr>
<td>$EK-L(\sigma_{EL} = \sigma_{KL})$</td>
<td>1.44 (1.73)</td>
<td>.502 (.742)</td>
</tr>
</tbody>
</table>

The first two have only eight quasi-concave points out of 464 observations while the last has only 85 such points. On the other hand, the unconstrained Translog and the equation which imposes $\sigma_{LE} = \sigma_{KE} = 1$ have only seven observations which violate the convexity conditions. The equation with $\sigma_{LK} = \sigma_{EK} \neq 1$ imposed has only ten non-convex points. Further, the absolute values of the bordered Hessians are very small ($10^{-7}$) and are not significantly different from zero. As the non-convex points are concentrated at the extremities of the observed output range it is not surprising that they violate the convexity conditions. For all the equations the total number of non-monotonic points ranged between 29 and 34, i.e., between 19 and 22 communities had one or more inputs with a computed negative marginal product. When the equations were re-estimated, with the offending observations deleted, there were no significant changes in the parameter estimates or $R^2$.

\[122\] All these tests are distributed as $F(2, 453)$. The notation $XY-Z$ means that $X$ and $Y$ are separable from $Z$. 
In sum, three separability hypotheses are consistent with the data and prior restrictions:

(i) $\sigma_{LE} = \sigma_{KE} = 1$
(ii) $\sigma_{LE} = \sigma_{KE} \neq 1$
(iii) $\sigma_{LK} = \sigma_{KE} \neq 1$

The three equations using (L1,E1,K3) are presented in columns (4), (5), and (6) of Table VI; the three equations using (L1,E1,K4) are presented in columns (7) to (9). Of course, only one of these conditions can hold. The difficulty may be that the direct estimation of the production function does not provide sufficient prior information to identify the parameters of the production function. The absence of constant returns precludes the use of the first-order conditions, unless they are modified as suggested in Chapter III. Those modified first-order conditions are estimated in Chapter VII, but unfortunately, the separability question remains unresolved.

In any event, the Translog equation with the hierarchy is accepted as the most useful specification of the urban area production function. On statistical grounds it is superior to the Cobb-Douglas and the hierarchy variable is highly significant; on pragmatic grounds, i.e., ease of interpretation and use, it is superior to the Translog which treats the hierarchy as an input.\(^{123}\) Since homogeneity is rejected for all the Translog equations, definitive statements on the degree of returns to scale are not possible. However, it appears that the hierarchy has a pronounced role in

\(^{123}\)This is not to say that this form of the Translog is useless, just that it is relatively difficult to use and interpret.
determining the degree of returns to scale.

(ii) Labour Adjusted for Human Capital (L2)

The results for the human capital adjusted labour variable are similar to those for L1. The Translog equations for (L2,E1,K3) and (L2,E1,K4), with the hierarchy, are presented in columns (1) and (2) of Table VII. In both equations the hierarchy variable is highly significant, the computed F(1,453)'s are 27.43 and 26.30 respectively. As well, the Cobb-Douglas specification is rejected at the 1% level for both equations (F(6,453)'s of 4.40 and 4.91 respectively). Again, homogeneity and linear homogeneity are rejected at the 1% level, F(3,453)'s of 4.18 and 4.05 and F(4,453)'s of 3.95 and 3.95 respectively, and returns to scale vary with the quantity of the urban area's inputs. As Exhibit C illustrates, the pattern for L2 is similar to that of L1. As one would expect (given the data in Tables II to IV), the logarithmic marginal product of L2 is greater than that of L1 while capital's logarithmic marginal product is lower. Otherwise the results are not sensitive to the input set used. At the median observation constant returns to physical inputs (t = 1.20) and increasing returns to physical inputs and the hierarchy (t = -4.86) prevail.

Exhibit C

Logarithmic Marginal Products Using L2, E1, and K3

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>E</th>
<th>K</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>G76</td>
<td>1.180</td>
<td>-.012</td>
<td>-.042</td>
<td>1.126</td>
</tr>
<tr>
<td>G5</td>
<td>.853</td>
<td>.011</td>
<td>.169</td>
<td>1.033</td>
</tr>
<tr>
<td>G4</td>
<td>.689</td>
<td>.023</td>
<td>.276</td>
<td>.988</td>
</tr>
<tr>
<td>G3</td>
<td>.578</td>
<td>.041</td>
<td>.336</td>
<td>.955</td>
</tr>
<tr>
<td>G2</td>
<td>.499</td>
<td>.044</td>
<td>.401</td>
<td>.944</td>
</tr>
<tr>
<td>G1</td>
<td>.423</td>
<td>.047</td>
<td>.469</td>
<td>.939</td>
</tr>
<tr>
<td></td>
<td>1.(L2,E1,K3)</td>
<td>2.(L2,E1,K4)</td>
<td>3.(L2,E1,K3)</td>
<td>4.(L2,E1,K3)</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td><strong>a_L</strong></td>
<td>.558499</td>
<td>.610277</td>
<td>.530559</td>
<td>.537047</td>
</tr>
<tr>
<td>(a_L)</td>
<td>(.0640367)</td>
<td>(.0759541)</td>
<td>(.0582796)</td>
<td>(.0601409)</td>
</tr>
<tr>
<td><strong>a_E</strong></td>
<td>.0370751</td>
<td>.0457701</td>
<td>.0498765</td>
<td>.0500429</td>
</tr>
<tr>
<td>(a_E)</td>
<td>(.034635)</td>
<td>(.0377011)</td>
<td>(.0309958)</td>
<td>(.0346656)</td>
</tr>
<tr>
<td><strong>a_K</strong></td>
<td>.356575</td>
<td>.283608</td>
<td>.377796</td>
<td>.373299</td>
</tr>
<tr>
<td>(a_K)</td>
<td>(.065454)</td>
<td>(.0790141)</td>
<td>(.0585489)</td>
<td>(.0593422)</td>
</tr>
<tr>
<td><strong>γL</strong></td>
<td>-.0154311</td>
<td>-.0937823</td>
<td>-.0208006</td>
<td>-.0167238</td>
</tr>
<tr>
<td>(γL)</td>
<td>(.088491)</td>
<td>(.0950534)</td>
<td>(.0881297)</td>
<td>(.0684288)</td>
</tr>
<tr>
<td><strong>γE</strong></td>
<td>.0317198</td>
<td>.0161718</td>
<td>-.0208536</td>
<td>-.0180031*</td>
</tr>
<tr>
<td>(γE)</td>
<td>(.0666075)</td>
<td>(.0669272)</td>
<td>(.0351285)</td>
<td>(.0351285)</td>
</tr>
<tr>
<td><strong>γK</strong></td>
<td>-.378937</td>
<td>-.479815</td>
<td>-.300330</td>
<td>-.311208</td>
</tr>
<tr>
<td>(γK)</td>
<td>(.137794)</td>
<td>(.147984)</td>
<td>(.0918622)</td>
<td>(.0917280)</td>
</tr>
<tr>
<td><strong>γE</strong></td>
<td>.0534363</td>
<td>.101287</td>
<td>0</td>
<td>.0180644*</td>
</tr>
<tr>
<td>(γE)</td>
<td>(.0884316)</td>
<td>(.0911004)</td>
<td>(.090984)</td>
<td>(.090984)</td>
</tr>
<tr>
<td><strong>γK</strong></td>
<td>.242679</td>
<td>.326325</td>
<td>.191002</td>
<td>.193863</td>
</tr>
<tr>
<td>(γK)</td>
<td>(.0932946)</td>
<td>(.105024)</td>
<td>(.0772815)</td>
<td>(.0731226)</td>
</tr>
<tr>
<td><strong>γE</strong></td>
<td>-.0926015</td>
<td>-.124982</td>
<td>0</td>
<td>-.0173517</td>
</tr>
<tr>
<td>(γE)</td>
<td>(.087617)</td>
<td>(.0930776)</td>
<td>(.0242733)</td>
<td>(.0242733)</td>
</tr>
<tr>
<td><strong>γK</strong></td>
<td>-.0640117</td>
<td>-.0146312</td>
<td>-.0586944</td>
<td>-.0579193</td>
</tr>
<tr>
<td>(γK)</td>
<td>(.0207216)</td>
<td>(.0216017)</td>
<td>(.0187225)</td>
<td>(.0186223)</td>
</tr>
<tr>
<td><strong>γH</strong></td>
<td>.184642</td>
<td>.169361</td>
<td>.179602</td>
<td>.179111</td>
</tr>
<tr>
<td>(γH)</td>
<td>(.0325744)</td>
<td>(.0329254)</td>
<td>(.0322757)</td>
<td>(.0318108)</td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>.9403</td>
<td>.9408</td>
<td>.9402</td>
<td>.9401</td>
</tr>
<tr>
<td><strong>SSR</strong></td>
<td>47.3968</td>
<td>47.0370</td>
<td>47.5451</td>
<td>47.5869</td>
</tr>
<tr>
<td><strong>DW</strong></td>
<td>2.06</td>
<td>2.08</td>
<td>2.06</td>
<td>2.06</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>464</td>
<td>464</td>
<td>464</td>
<td>464</td>
</tr>
</tbody>
</table>

\(\sigma_{LE}^{a,KE}\) \(\sigma_{LK}^{a,KE}\) \(\sigma_{LE}^{a,KE}\) \(\sigma_{LK}^{a,KE}\)
When the question of separability is considered the results in Exhibit D are obtained.

Exhibit D

(The numbers in brackets refer to the input set L2, E1, and K4; the other F-tests refer to L2, E1, and K3)

<table>
<thead>
<tr>
<th></th>
<th>Linear ($\sigma_{ij} = 1$)</th>
<th>Non-linear ($\sigma_{ij} \neq 1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LK-E($\sigma_{EL} = \sigma_{KE}$)</td>
<td>.709 (.91)</td>
<td>4.22** (4.29**)</td>
</tr>
<tr>
<td>LE-K($\sigma_{KL} = \sigma_{KE}$)</td>
<td>2.92* (5.06**)</td>
<td>.908 (1.11)</td>
</tr>
<tr>
<td>EK-L($\sigma_{LE} = \sigma_{LK}$)</td>
<td>2.80* (5.06**)</td>
<td>4.91** (7.32**)</td>
</tr>
</tbody>
</table>

*significant at the 5% level
**significant at the 1% level

In this case, all the separability hypotheses, except for $\sigma_{EK} = \sigma_{EL} = 1$ and $\sigma_{KL} = \sigma_{KE} \neq 1$, are rejected at the 5% level for (L2, E1, K3) and at the 1% level for (L2, E1, K4). The two acceptable equations for each variable set are presented in columns (3) to (6) of Table VII. The equations with L2 exhibit positive marginal products but the second order conditions are invariably violated. However, since the point estimates which are used to evaluate the bordered Hessian have large standard errors, this is not a serious problem. For the sake of comparison, the Cobb-Douglas equations for (L1, E1, K3) and (L2, E1, K4) are presented in columns (7) and (8) of Table VII.

---

124 As before, all these tests are distributed as $F(2,453)$.

125 For the Cobb-Douglas using (L2, E1, K3) and the hierarchy, increasing returns to physical inputs are accepted ($t = -2.23$), while for (L2, E1, K4) and the hierarchy, constant returns to scale cannot be rejected ($t = -1.58$).
In summary, when labour unadjusted for human capital is used in the production function, returns to scale are variable and increase with movements up the urban hierarchy. When labour is adjusted for human capital the same pattern emerges. Thus, the results with Translog, using L1 and L2, contradict the assumptions of the urban models discussed in Part I of Chapter II; the results are, however, consistent with hierarchy models of urban centres. Because of the violation of the second order conditions, one could choose to reject the Translog with L2 and to accept the Cobb-Douglas equation.\textsuperscript{126} In this case, the results are consistent with the urban models of Conrod, Mills, Dixit, etc., but it is also clear that, even in this case, the hierarchy is an important factor in determining the extent of returns to scale in the urban area. After evaluating the other possibilities for implementing the hierarchy a more complete discussion of the implications of these results will be given. Finally, given the results of the separability tests using L2, it is reasonable to reduce the set of acceptable separability hypotheses to two:

\[(i)\] \(\sigma_{EL} = \sigma_{EK} = 1\)

\[(ii)\] \(\sigma_{KL} = \sigma_{KE} \neq 1\).

C. The Continuous-Discrete Hierarchy\textsuperscript{127}

The Continuous-Discrete Hierarchy assumes that the parameters of the

\textsuperscript{126}This, of course, also requires accepting the hypotheses of complete separability, i.e., \(\sigma_{KE} = \sigma_{KL} = \sigma_{LE} = 1\), of constant factor shares, and that marginal products are proportional to average products.

\textsuperscript{127}Since K3 and K4 consistently generate similar results, only the estimates using K3 will be reported in the remainder of this study.
production function are identical for each hierarchy level except for the constant. The implementation proceeds by using 0-1 dummies for each hierarchy level with the median hierarchy level (where Harris is also located) used as the base. Using L1, when the six dummy variables are added to the Cobb-Douglas equation the \( F(6,454) \) is 1.03; when added to the Translog the \( F(6,448) \) is 1.05. When L2 is used the Cobb-Douglas and Translog F-statistics are, respectively, 1.40 and 1.21. The continuous-discrete hierarchy does not increase the explanatory power of the model and is rejected.

D. The Discrete Hierarchy

The implementation of the discrete hierarchy requires estimating the production function for each group. Since there are only two observations in Group 7 it is merged with Group 6. The Translog function is estimated for each group and the special case of a Cobb-Douglas technology is accepted for a group if the SSR is not significantly increased. When L1 is used, the Cobb-Douglas technology is rejected for G4 and G1 with F-statistics of 5.30 and 2.36 distributed as \((6,66)\) and \((6,94)\), the former being significant at the 1% level and the latter at the 5% level. For L2, the Cobb-Douglas technology is rejected for G4 at the 5% level with an \( F(6,66) \) of 2.22.

The total of the SSR of the six accepted equations is then compared to the SSR of the Cobb-Douglas and Translog equations without a hierarchy variable, for the entire sample.\(^{128}\)

\(^{128}\)This is not, perhaps, the most obvious procedure. If the SSR for the Translog for the entire sample is compared to the sum of the SSR for the Translog estimated for each of the six groups, the computed \( F(50,404) \) is .924 and .505 for L1 and L2 respectively. If this procedure is followed for the Cobb-Douglas, the \( F(20,440) \)'s for L1 and L2 are 1.11 and .893. None
The F-statistics are:

<table>
<thead>
<tr>
<th></th>
<th>Cobb-Douglas (dof)</th>
<th>Translog (dof)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>1.53* (32,428)</td>
<td>1.44 (26,428)</td>
</tr>
<tr>
<td>L2</td>
<td>.83 (26,434)</td>
<td>.04 (20,434)</td>
</tr>
</tbody>
</table>

The group equations using L1 are significant at the 5% level (the critical value is 1.47) when compared to the Cobb-Douglas, and are almost significant when compared to the Translog (the critical value is 1.53). The groups with L2 are not significant. Testing the three accepted Translog equations for linear homogeneity results in its acceptance for L1's groups G4 and G1, with computed F-statistics of .468 and 1.20 respectively, and for L2's group G4, F(4,66) = .386. The constrained within group estimates for L1 and L2 are presented in Tables VIII and IX. However, the rest of this section's discussion is confined to Table VIII.

There are two possible explanations for the acceptance of homogeneity for all of the group equations while it is rejected for the equations using the entire sample. Firstly, homogeneity can be incorrectly rejected if factor prices vary between communities. It was argued previously that prices should not vary substantially but, given the wide range of output and input levels, a small variance in prices could result in homogeneity being rejected for the entire sample but accepted within groups. However, the more
<table>
<thead>
<tr>
<th></th>
<th>L1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G7-6</td>
<td>G5</td>
<td>G4</td>
<td>G3</td>
<td>G2</td>
<td>G1</td>
</tr>
<tr>
<td>aL</td>
<td>1.04570</td>
<td>.796071</td>
<td>.513043</td>
<td>.380049</td>
<td>.426970</td>
<td>.426539</td>
</tr>
<tr>
<td></td>
<td>(.0613533)</td>
<td>(.243361)</td>
<td>(.313686)</td>
<td>(.106676)</td>
<td>(.0984557)</td>
<td>(.155038)</td>
</tr>
<tr>
<td>aE</td>
<td>-.0631958*</td>
<td>.082435</td>
<td>.159696*</td>
<td>.0960125</td>
<td>-.059042*</td>
<td>-.069857*</td>
</tr>
<tr>
<td></td>
<td>(.33673)</td>
<td>(.058762)</td>
<td>(.06113)</td>
<td>(.0521017)</td>
<td>(.055876)</td>
<td>(.0954594)</td>
</tr>
<tr>
<td>aK</td>
<td>.0174958</td>
<td>.212221</td>
<td>.327251</td>
<td>.667212</td>
<td>.632072</td>
<td>.643318</td>
</tr>
<tr>
<td></td>
<td>(.071433)</td>
<td>(.251213)</td>
<td>(.315511)</td>
<td>(.122900)</td>
<td>(.106531)</td>
<td>(.206874)</td>
</tr>
<tr>
<td>γLL</td>
<td>.410635</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.01333</td>
</tr>
<tr>
<td></td>
<td>(.882637)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.337323)</td>
</tr>
<tr>
<td>γEE</td>
<td>.001824*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.0066448</td>
</tr>
<tr>
<td></td>
<td>(1.4378)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.78588)</td>
</tr>
<tr>
<td>γKK</td>
<td>-.564987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.89432</td>
</tr>
<tr>
<td></td>
<td>(.833235)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.40090)</td>
</tr>
<tr>
<td>γKL</td>
<td>.078088*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.957147*</td>
</tr>
<tr>
<td></td>
<td>(.849196)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.3312245)</td>
</tr>
<tr>
<td>γLE</td>
<td>-.488723</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.0561826</td>
</tr>
<tr>
<td></td>
<td>(.160111)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.148879)</td>
</tr>
<tr>
<td>γKE</td>
<td>.446899*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.0626274*</td>
</tr>
<tr>
<td></td>
<td>(1.45222)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.715411)</td>
</tr>
<tr>
<td>C</td>
<td>-.261223</td>
<td>-.315339</td>
<td>-.228701</td>
<td>-.287118</td>
<td>-.241474</td>
<td>-.286389</td>
</tr>
<tr>
<td></td>
<td>(.0391978)</td>
<td>(.0983396)</td>
<td>(.0555994)</td>
<td>(.040744)</td>
<td>(.0508626)</td>
<td>(.0674266)</td>
</tr>
<tr>
<td>R²</td>
<td>.9919</td>
<td>.9668</td>
<td>.9282</td>
<td>.8532</td>
<td>.8029</td>
<td>.8007</td>
</tr>
<tr>
<td>DW</td>
<td>2.07</td>
<td>2.02</td>
<td>1.80</td>
<td>1.58</td>
<td>1.24</td>
<td>1.66</td>
</tr>
<tr>
<td>N</td>
<td>14</td>
<td>35</td>
<td>76</td>
<td>104</td>
<td>92</td>
<td>143</td>
</tr>
<tr>
<td>SSR</td>
<td>.0465489</td>
<td>.886986</td>
<td>2.29922</td>
<td>9.83275</td>
<td>10.8826</td>
<td>24.0263</td>
</tr>
<tr>
<td>L2</td>
<td>G7-6</td>
<td>G5</td>
<td>G4</td>
<td>G3</td>
<td>G2</td>
<td>G1</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>$\alpha_L$</td>
<td>.944518</td>
<td>.511725</td>
<td>.633252</td>
<td>.501117</td>
<td>.568189</td>
<td>.400230</td>
</tr>
<tr>
<td></td>
<td>(.072643)</td>
<td>(.261880)</td>
<td>(.131920)</td>
<td>(.108870)</td>
<td>(.076215)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_E$</td>
<td>-.029384</td>
<td>.095484</td>
<td>-.00952*</td>
<td>.0704945</td>
<td>-.0010653</td>
<td>.0580817</td>
</tr>
<tr>
<td></td>
<td>(.0545358)</td>
<td>(.062383)</td>
<td>(.037237)</td>
<td>(.0511444)</td>
<td>(.0750904)</td>
<td>(.0734513)</td>
</tr>
<tr>
<td>$\alpha_K$</td>
<td>.143989</td>
<td>.521846</td>
<td>.376268</td>
<td>.589582</td>
<td>.569784</td>
<td>.754358</td>
</tr>
<tr>
<td></td>
<td>(.0799848)</td>
<td>(.257745)</td>
<td>(.127386)</td>
<td>(.115862)</td>
<td>(.111406)</td>
<td>(.0969445)</td>
</tr>
<tr>
<td>$\gamma_{LL}$</td>
<td>1.88517</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.0979)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{EE}$</td>
<td></td>
<td>-.021129*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.9957)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{KK}$</td>
<td></td>
<td>.449597</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.996595)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{KL}$</td>
<td></td>
<td>-1.17948*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.0368)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{LE}$</td>
<td></td>
<td>-.707222</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.204235)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{KE}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.728351*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.0237)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td>-.00686848</td>
<td>-.0997488</td>
<td>-.0508581</td>
<td>-.179563</td>
<td>-.0897678</td>
<td>-.0188479</td>
</tr>
<tr>
<td></td>
<td>(.0826824)</td>
<td>(.0872203)</td>
<td>(.0309457)</td>
<td>(.0320475)</td>
<td>(.0368946)</td>
<td>(.0541404)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.9974</td>
<td>.9600</td>
<td>.9289</td>
<td>.8630</td>
<td>.8644</td>
<td>.7202</td>
</tr>
<tr>
<td>DW</td>
<td>2.39</td>
<td>1.72</td>
<td>1.39</td>
<td>1.51</td>
<td>1.50</td>
<td>1.40</td>
</tr>
<tr>
<td>N</td>
<td>14</td>
<td>35</td>
<td>76</td>
<td>104</td>
<td>92</td>
<td>143</td>
</tr>
<tr>
<td>SSR</td>
<td>.060692</td>
<td>1.06231</td>
<td>2.27633</td>
<td>9.17769</td>
<td>9.40697</td>
<td>28.6987</td>
</tr>
</tbody>
</table>
reasonable explanation for the acceptance of different hypotheses for the
group equations is that different types of goods are produced and supplied
by each hierarchy level. Thus, the production functions for each hierarchy
level differ.\textsuperscript{129}

The obvious item of interest is the changing pattern of coefficients.
Starting from the lowest level (G1), there is a steady rise in $\alpha_L$, offset,
in part, by a decline in $\alpha_K$. The parameter on land reaches a peak in the
middle hierarchy level.\textsuperscript{130} Thus, it would appear that labour's productivity
and share of output (income) increases with movements up the hierarchy, i.e.,
with urbanization.\textsuperscript{131} More precision on this point may be possible after
estimating the factor share equations in Chapter VII.

A second issue concerns the existence of DRS, CRS, and IRS within
each group. The results for each group are:

<table>
<thead>
<tr>
<th>t-statistic (dof)</th>
<th>Accepted Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>G76   -1.38 (10)</td>
<td>CRS</td>
</tr>
<tr>
<td>G5    -2.18 (31)</td>
<td>IRS (at 5% level)</td>
</tr>
<tr>
<td>G4    .361 (65)</td>
<td>CRS</td>
</tr>
<tr>
<td>G3    -2.61 (100)</td>
<td>IRS</td>
</tr>
<tr>
<td>G2    -1.92 (88)</td>
<td>CRS (5% critical value is 2.00)</td>
</tr>
<tr>
<td>G1    -.301 (134)</td>
<td>CRS</td>
</tr>
</tbody>
</table>

In contrast to the continuous hierarchy approach, increasing returns to
physical inputs are observed in two hierarchy levels and decreasing returns

\textsuperscript{129}This is an argument for accepting the group equations even if they
are statistically insignificant.

\textsuperscript{130}Though it is of some concern that $\alpha_E$ is occasionally negative it is,
in such cases, insignificantly different from zero.

\textsuperscript{131}In part this is due to a greater concentration of labour intensive
activities (services) in larger centres.
in none. Thus, it would appear that G5 and G3 are the levels where continued
growth is possible without government intervention. For the other groups
there is no natural incentive to growth but they can be influenced by chance
or government intervention. Of course, if the system is left undisturbed
then there will be a tendency for G5 and G3 communities to grow faster than
the others and, given a declining or constant population, this means negative
growth for communities in other groups. But this negative growth is not
necessarily because a community is non-viable, but because other communities
are more attractive. It is phenomena such as this which explain the gradual
reduction of the number of hierarchy levels and communities in Saskatchewan,132
phenomena neglected in most urban models.

E. Summary and Implications

One of the major purposes of this study is to evaluate the consistency
of the postulates and structure of urban models with respect to empirical
observation. While the basic assumption of increasing returns remains
unscathed, this study finds that factors external to the urban area play an
important role. Most urban models have sought the causes of urban growth
and size within the city, it should now be clear that such a simple approach
is inappropriate.

In the case of L1 it appears that increasing returns to physical inputs
result from omitting a hierarchy variable from the production function.
However, the urban models discussed in Chapter II invariably rule out external
influences and assume increasing returns to physical inputs within individual

132 The observation that Saskatchewan is gradually developing a three-
level hierarchy is suggested in Stabler and Williams (1972).
urban areas. If that assumption is incorrect, if factors external to an individual urban area cause increasing returns, then theoretical results (especially about the optimum city) and policy recommendations based on that assumption are likely in error. But things are never that clearcut. When L2 is used with the hierarchy, increasing returns prevail for the Cobb-Douglas and variable returns for the Translog. This is significant since the necessary and sufficient condition for generating urban areas and obtaining the static results of the standard urban model is increasing returns to physical inputs within the urban area. But since the Translog with L2 is not quasi-concave the hypothesis of increasing returns to physical inputs cannot really be rejected. What is clear is that returns to scale vary with urban size and hierarchy position. This suggests an important omission from most urban models: urban areas are phenomena of the hierarchy, they result from specialization in goods with different scale economies, and these factors must be included in a model of the urban area.

There are, of course, replies which can be made in defence of the standard model. The most obvious is that the hierarchy variable simply embodies the postulated agglomeration economies, but this does not refute the basic finding, that increasing returns to scale are external to the urban area. One might argue that the hierarchy variable is only a proxy for the changing mix of activities and results in understating the degree

---

133 Abstracting, as always, from locational and non-economic factors.

134 This is implicit in Conrod (1967) and explicit in Evans (1972) and Henderson (1974).

135 See Chapter III.
of returns to physical inputs. Aside from noting that the causes of the changing mix are external to individual urban areas and are not considered in most urban models, the results for the within group estimations refute this objection. The groups are structurally similar and a test of the hypothesis of IRS for each group is as fair a test of the standard model as is possible. In four out of six groups CRS is accepted, while IRS is accepted for just two. Of course, it must be noted that the acceptance of constant returns to scale may also, but not necessarily, be inconsistent with most hierarchy models.

Of some surprise, especially in light of earlier remarks, is the relatively poor performance of the various forms of the discrete hierarchy. However, the results for these approaches depend crucially on the correctness of the clustering algorithm, its unconstrained use, and the completeness of the descriptive data.

Another unresolved problem is which of the separability hypotheses should be adopted.\textsuperscript{136} If $\sigma_{KE} = \sigma_{LE}$ then capital and labour are separable from land, which implies that higher land taxes, for example, would not alter the ratio of the marginal products for capital and labour in an urban area. On the other hand, if $\sigma_{LK} = \sigma_{KE}$ then land and labour are separable from capital and higher land taxes will affect the ratio of the marginal products for capital and labour.

A final question is who gets the extra output associated with movements up the hierarchy? Decreasing returns mean that at least one factor is getting more than its marginal product and increasing returns mean that some factors

\textsuperscript{136} Of course, the data do not allow the selection of one or the other on statistical grounds.
are getting less than their marginal products. Of course, it is possible that various combinations of "more" and "less" prevail. There should be no predisposition, in the absence of CRS, to believe that the market will decide that the marginal product of a factor is the minimum (if DRS prevail) or maximum (if IRS prevail) that factor can receive.\textsuperscript{137}

The next chapter attempts to answer the income distribution questions associated with IRS, DRS, and urbanization.

III. SUMMARY AND CONCLUSIONS

This chapter was concerned with estimating the parameters of the urban area production function and with determining the role and importance of the urban hierarchy. Generally, the Translog production function with a hierarchy variable yielded results consistent with prior expectations and technical restrictions. The finding of decreasing or constant returns to physical inputs (when the hierarchy is included in the equation) indicates a re-evaluation of existing urban models is required. The importance of the hierarchy is emphasized by the results which show that factor marginal products vary with the urban hierarchy. Both the continuous and discrete hierarchies were statistically significant, the former at the 1\% level and the latter at the 5\% level.

In conclusion, it has proved possible to estimate an aggregate production function for the urban area. In this respect recent models of the urban are supported, though it is clear that the urban hierarchy cannot be ignored.

\textsuperscript{137} Even when CRS prevails there does not seem to be any compelling reason for factors to receive their marginal products. Market imperfections can exist under any technological regime.
Chapter VII

ESTIMATION OF THE COST-SHARE EQUATIONS

I. INTRODUCTION

In this chapter the cost-share equations of Chapter III, equations (14'), are estimated. The production function and the cost-share equations for labour and capital are stacked and estimated by iterating the Seemingly Unrelated Regression Equations estimator (ISURE). As indicated in Chapter IV, the cost-shares are measured by each factor's share of total income. Unfortunately, while aggregate income is reasonably accurate, the income share data for most of the communities (especially the smallest) is unreliable. This problem is due to the categorization and residency problems discussed in Chapter IV. The major difficulty is that land and capital shares are underreported for the small communities. Accordingly, the cost-share estimation has been separated from the direct estimation of the production function in Chapter VI.

Despite the data problems, a consistent pattern emerges. First, in accordance with the forces of equilibrium, there is a tendency for factor rewards to be equalized between urban centres; second, for the entire sample, there is a transfer of income from capital to labour in the sense that capital receives less than its marginal product while labour receives more; finally, except in the case of one hierarchy level, when the c_i's within the hierarchy level are not equal to zero, there is a transfer of income from capital to labour. Section II discusses these results while Section III

138 Many communities report zero earned income for capital and land.
concludes the chapter.

II. COST-SHARE EQUATIONS

The cost-share equations are estimated for the entire sample and for each of the groups of Chapter V, though computer limitations prevent a complete analysis of the entire sample. To minimize the problems with the data and to make use of prior information, the production function is estimated with the cost-share equations. However, this procedure is not completely successful. The data for land shares, and to a lesser extent for capital shares, are poor and it proved impossible to obtain sensible results for the land equations using the entire sample. The land equation is also statistically insignificant when estimated by ordinary least squares (OLS), i.e., the $R^2$ is not significantly different from zero. Finally, computer limitations prevent the testing of the separability hypotheses.

Estimating each cost-share equation by OLS does not allow the imposition of symmetry, permit tests of homogeneity, or tests of the degree of returns to scale. It is therefore necessary to stack the equations. However, the residual terms of the three cost-share equations must sum to zero so that only two of the cost-share equations can be estimated with the production function. (The third equation is determined residually.) But with one of the equations deleted both the stacked OLS and the SURE estimator give biased estimates. However, the iterated SURE estimator gives unbiased estimates which are independent of the omitted equation.\(^\text{139}\) Accordingly, the stacked equation system consisting of the production function and two cost-share equations is estimated with the ISURE estimator.

\(^{139}\)Kmenta and Gilbert (1968).
The cost-share equations are first estimated by OLS for the entire sample. While the equation for land is insignificant, the equations for labour and capital are significant at the 5% level, \( F(3,460) = 3.16 \), and at the 1% level, \( F(4,450) = 4.28 \), respectively. These two equations are presented in Table X, column (1).\(^{140}\) Only \( c_K \) is significant. If \( \alpha_E \) is set equal to its value in column (1) of Table VI, the \( \alpha_i \)'s sum to 1.006 and the sign of \( c_K \) is consistent with prior expectations.

Imposing symmetry on the labour and capital equations resulted in an insignificant loss of explanatory power (\( F = .49 \)) and, using the ISURE estimator, both \( c_L \) and \( c_K \) are significantly different from zero (Table X, column (2)). Again, following the previous procedure for \( \alpha_E \), the \( \alpha_i \)'s sum to 1.002. Thus, at least one \( c_i \) must be less than zero--in this case, \( c_K \). The sign of \( c_L \) is consistent with the pattern of factor marginal products noted in Chapter VI, section II-B. This result indicates that the assumption of \((n-1)\) factors receiving their marginal products while the \( \bar{n} \)th receives the residual is another oversimplification of the standard model--though not a serious one. Finally, since \( c_E \) is not significantly different from zero in the OLS estimation, it is reasonable to conclude that \( c_L \) and \( c_K \) reflect the manner in which increases in output due to the hierarchy are distributed. Accordingly, in the remainder of this chapter, \( c_E \) and land's share are determined residually.\(^{141}\)

For the within group estimation the procedure is to:

\(^{140}\) There are two estimates of \( \gamma_{KL} \) because symmetry has not been imposed.

\(^{141}\) A priori it is reasonable to expect the most immobile factor (land) to bear all the residual income adjustments after labour and capital are paid.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_L$</td>
<td>0.775080</td>
<td>.764353</td>
<td>.947418</td>
<td>.810628</td>
<td>.769364</td>
<td>.777825</td>
<td>.287064</td>
<td>.744829</td>
<td>.531157</td>
</tr>
<tr>
<td></td>
<td>(.00630049)</td>
<td>(.0084411)</td>
<td>(.0211758)</td>
<td>(.0069802)</td>
<td>(.00614616)</td>
<td>(.00802562)</td>
<td>(.0301267)</td>
<td>(.0150839)</td>
<td>(.0280116)</td>
</tr>
<tr>
<td>$a_E$</td>
<td>-</td>
<td>-</td>
<td>.0117869</td>
<td>.042931*</td>
<td>.045615*</td>
<td>.042615*</td>
<td>.132168</td>
<td>- .0863476</td>
<td>.0694641</td>
</tr>
<tr>
<td></td>
<td>(.0315946)</td>
<td>(.0143316)</td>
<td>(.0113398)</td>
<td>(.0156659)</td>
<td>(.0565459)</td>
<td>(.0568540)</td>
<td>(.0515543)</td>
<td>(.0515543)</td>
<td>(.0515543)</td>
</tr>
<tr>
<td>$a_K$</td>
<td>.199078</td>
<td>.205905</td>
<td>.104615</td>
<td>.146441</td>
<td>.185021</td>
<td>.179560</td>
<td>.676139</td>
<td>.216638</td>
<td>.468148</td>
</tr>
<tr>
<td></td>
<td>(.01345514)</td>
<td>(.00882195)</td>
<td>(.0070391)</td>
<td>(.00752576)</td>
<td>(.00550492)</td>
<td>(.00802852)</td>
<td>(.0124644)</td>
<td>(.0284948)</td>
<td>(.0284948)</td>
</tr>
<tr>
<td>$\gamma_{LL}$</td>
<td>.0180775</td>
<td>.0101566</td>
<td>.0103902</td>
<td>.0137474</td>
<td>.012624</td>
<td>.019963</td>
<td>.0263361</td>
<td>.0500026</td>
<td>.0507128</td>
</tr>
<tr>
<td></td>
<td>(.013482195)</td>
<td>(.0070391)</td>
<td>(.00752576)</td>
<td>(.00550492)</td>
<td>(.00802852)</td>
<td>(.0124644)</td>
<td>(.0284948)</td>
<td>(.0284948)</td>
<td>(.0284948)</td>
</tr>
<tr>
<td>$\gamma_{EE}$</td>
<td>-</td>
<td>-</td>
<td>.0472146</td>
<td>.0730396</td>
<td>.121794</td>
<td>.0296398</td>
<td>.208227</td>
<td>.0308845</td>
<td>.0308845</td>
</tr>
<tr>
<td></td>
<td>(.013482195)</td>
<td>(.0070391)</td>
<td>(.00752576)</td>
<td>(.00550492)</td>
<td>(.00802852)</td>
<td>(.0124644)</td>
<td>(.0284948)</td>
<td>(.0284948)</td>
<td>(.0284948)</td>
</tr>
<tr>
<td>$\gamma_{KK}$</td>
<td>.0193341</td>
<td>.0208993</td>
<td>.0145668</td>
<td>.0141632</td>
<td>.0193341</td>
<td>.0208993</td>
<td>.0145668</td>
<td>.0141632</td>
<td>.0145668</td>
</tr>
<tr>
<td></td>
<td>(.0145668)</td>
<td>(.0141632)</td>
<td>(.0145668)</td>
<td>(.0141632)</td>
<td>(.0145668)</td>
<td>(.0141632)</td>
<td>(.0145668)</td>
<td>(.0141632)</td>
<td>(.0145668)</td>
</tr>
<tr>
<td>$\gamma_{KE}$</td>
<td>- .0236558</td>
<td>- .0215185</td>
<td>- .0071856</td>
<td>- .00660417</td>
<td>- .0236558</td>
<td>- .0215185</td>
<td>- .0071856</td>
<td>- .00660417</td>
<td>- .0236558</td>
</tr>
<tr>
<td></td>
<td>(.0071856)</td>
<td>(.00660417)</td>
<td>(.0071856)</td>
<td>(.00660417)</td>
<td>(.0071856)</td>
<td>(.00660417)</td>
<td>(.0071856)</td>
<td>(.00660417)</td>
<td>(.0071856)</td>
</tr>
<tr>
<td>$\gamma_{KL}$</td>
<td>- .0193341</td>
<td>- .0144678</td>
<td>- .0143365</td>
<td>- .0132331</td>
<td>- .112449</td>
<td>- .0285660</td>
<td>- .205321</td>
<td>- .0270598</td>
<td>- .055658</td>
</tr>
<tr>
<td></td>
<td>(.0143365)</td>
<td>(.0132331)</td>
<td>(.0143365)</td>
<td>(.0132331)</td>
<td>(.0143365)</td>
<td>(.0132331)</td>
<td>(.0143365)</td>
<td>(.0132331)</td>
<td>(.0143365)</td>
</tr>
<tr>
<td>$\gamma_{LE}$</td>
<td>.0185395</td>
<td>.0163870</td>
<td>.01875045</td>
<td>(.00718427)</td>
<td>.0516360</td>
<td>(.039500)</td>
<td>.0359128</td>
<td>.035568</td>
<td>.0359800</td>
</tr>
<tr>
<td></td>
<td>(.00718427)</td>
<td>(.00718427)</td>
<td>(.00718427)</td>
<td>(.00718427)</td>
<td>(.00718427)</td>
<td>(.00718427)</td>
<td>(.00718427)</td>
<td>(.00718427)</td>
<td>(.00718427)</td>
</tr>
<tr>
<td>C</td>
<td>-</td>
<td>-</td>
<td>- .420532</td>
<td>- .172734</td>
<td>- .280098</td>
<td>- .358225</td>
<td>- .390923</td>
<td>- .344322</td>
<td>- .342765</td>
</tr>
<tr>
<td></td>
<td>(.0547292)</td>
<td>(.0292522)</td>
<td>(.0216929)</td>
<td>(.0328629)</td>
<td>(.0338263)</td>
<td>(.0358361)</td>
<td>(.0329039)</td>
<td>(.0329039)</td>
<td>(.0329039)</td>
</tr>
<tr>
<td>$C_L$</td>
<td>0</td>
<td>.0083996</td>
<td>- .0943472</td>
<td>.0000000</td>
<td>.0000000</td>
<td>.0000000</td>
<td>.0000000</td>
<td>.0000000</td>
<td>.0000000</td>
</tr>
<tr>
<td></td>
<td>(.00353203)</td>
<td>(.0167739)</td>
<td>(.0167739)</td>
<td>(.0167739)</td>
<td>(.0167739)</td>
<td>(.0167739)</td>
<td>(.0167739)</td>
<td>(.0167739)</td>
<td>(.0167739)</td>
</tr>
<tr>
<td>$C_K$</td>
<td>- .0195322</td>
<td>- .2444565</td>
<td>.0000000</td>
<td>.0000000</td>
<td>.0000000</td>
<td>.0000000</td>
<td>.0000000</td>
<td>.0000000</td>
<td>.0000000</td>
</tr>
<tr>
<td></td>
<td>(.00868529)</td>
<td>(.00470088)</td>
<td>(.00470088)</td>
<td>(.00470088)</td>
<td>(.00470088)</td>
<td>(.00470088)</td>
<td>(.00470088)</td>
<td>(.00470088)</td>
<td>(.00470088)</td>
</tr>
<tr>
<td>$\bar{R}^2$</td>
<td>ML .0202</td>
<td>ML .0170</td>
<td>.9977</td>
<td>.9605</td>
<td>.8756</td>
<td>.8274</td>
<td>.8389</td>
<td>.7437</td>
<td>.8475</td>
</tr>
<tr>
<td></td>
<td>MK .0359</td>
<td>MK .0350</td>
<td>.9977</td>
<td>.9605</td>
<td>.8756</td>
<td>.8274</td>
<td>.8389</td>
<td>.7437</td>
<td>.8475</td>
</tr>
<tr>
<td>DM</td>
<td>ML 2.06</td>
<td>ML 2.07</td>
<td>2.20</td>
<td>1.56</td>
<td>1.97</td>
<td>1.62</td>
<td>1.20</td>
<td>1.37</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td>MK 2.04</td>
<td>MK 2.04</td>
<td>2.20</td>
<td>1.56</td>
<td>1.97</td>
<td>1.62</td>
<td>1.20</td>
<td>1.37</td>
<td>1.60</td>
</tr>
<tr>
<td>N</td>
<td>464</td>
<td>464</td>
<td>14</td>
<td>35</td>
<td>76</td>
<td>104</td>
<td>92</td>
<td>144</td>
<td>104</td>
</tr>
</tbody>
</table>
(1) estimate the stacked Translog production function and cost-share equations,
(2) test for the acceptability of the Cobb-Douglas, and
(3) test for homogeneity and linear homogeneity.

As indicated above, the cost-share equation for land is the omitted equation.

The parameter estimates of the cost-share equations differ substantially from those obtained in Chapter VI, section II-D. The obvious difference is that the Cobb-Douglas can only be rejected for G2 (versus G4 and G1 in Chapter VI), but of more interest is the difference in the accepted hypotheses regarding returns to scale. The results are presented in Table X, columns (3) to (8). For G76 the Cobb-Douglas cannot be rejected, $F(6,4) = .019$,\(^{142}\) but now increasing returns to scale are accepted, $t = -3.68$. This implies at least one $c_i \neq 0$. In this case, $c_K = 0$ and $c_L < 0$, which is not surprising if there is a tendency to factor payment equalization since the marginal product of labour in G76 is relatively high.\(^{143}\) For G5 the Cobb-Douglas and constant returns to scale are accepted, $F(6,25) = .237$ and $t = -1.67$. However, if constant returns are not imposed and the $c_i$ parameters are estimated, some are significantly different from zero. This happens again in G4. This result could be spurious since the cost-share data are unreliable; it could be due to adjustment lags; or, it could suggest that factors do not receive their marginal products even when constant returns prevail. All things considered it would seem that one of the first

\(^{142}\) In all cases this statement means that the Translog specification does not reduce the sum of squared residuals significantly.

\(^{143}\) For further confirmation see Exhibits A and C of Chapter VI.
two hypotheses, or some combination thereof, is the more likely.

The results for G4 deviate from the group results with just the production function as the Cobb-Douglas cannot be rejected, $F(6,66) = .367$, though constant returns are accepted, $t = - .120$. For G3 the Cobb-Douglas is again consistent with the data, $F(6,94) = 1.79$, and constant returns are also accepted, $t = 0.973$. As well, all the $c_i$'s are insignificant.

For G2 the Translog is accepted at the 5% level, $F(6,92) = 2.22$. The interesting aspect of G2 is the low $\hat{a}_L$ and high $\hat{a}_K$. Both linear homogeneity and homogeneity are rejected at the 1% level, $F(4,82) = 3.65$ and $F(3,82) = 3.99$, and, therefore, returns to scale vary. The attainment of factor payments consistent with those in other hierarchy levels is assured by the large positive $c_L$ and the large negative $c_K$. Again, this is consistent with earlier observations. The large $c_i$'s and the rejection of homogeneity would seem to suggest that this hierarchy level is in a state of disequilibrium, that it is unstable, or that it is a transitional hierarchy level. Similar results and patterns exist for the G3 Translog (Table X, column (9)). 144

The Cobb-Douglas cannot be rejected for GI, $F(6,134) = .48$, and the hypothesis of decreasing returns is "on the line," $t = 1.95$. Since both $c_L$ and $c_K$ are significant it seems reasonable to accept decreasing returns to scale. Of course, the acceptance of that hypothesis implies that GI and many of its communities are non-viable or, at the least, poor prospective growth centres.

144 Linear homogeneity is rejected at the 5% level, $F(4,94) = 3.19$, and homogeneity is rejected at the 1% level, $F(3,94) = 3.84$. As well, labour receives more ($c_L > 0$) and capital less ($c_K < 0$) than their marginal products.
III. CONCLUSIONS

Despite the limitations of the data it has been possible to determine how income is distributed when non-constant returns to scale prevail. While the income shift is from capital to labour for the entire sample, there is some diversity between hierarchy levels. Given the limited importance and immobility of land and the mobility of labour relative to capital, it is not surprising that capital bears the brunt of most of the negative adjustments. The cost-share equations, aside from the further demonstration that labour's productivity rises with movements up the hierarchy, seem to indicate that labour is the major beneficiary of the process of urbanization. Finally, the neo-classical distribution hypothesis is verified for technological regimes which exhibit constant returns to scale.

In conclusion, more attention must be given to specifying and justifying income distributional assumption when non-constant returns prevail. Institutional factors must be playing a role and could well have more than just distributional effects.
Chapter VIII

SUMMARY AND CONCLUSIONS

At the beginning of this study it was stated that the objective was to implement empirically recent theoretical models of the urban area. The hope was to determine if such an approach was feasible, if the basic assumptions and structure of the model were correct, and, given non-constant returns, how the income distribution problem could be resolved. A final objective was to unite the theoretical model with the empirical reality of the urban hierarchy. Chapter II outlined the theoretical models indicating which assumptions were testable hypotheses and empirical work bearing on the model's validity was cited. A similar discussion of the urban hierarchy was also included. Chapter III outlined the model in general terms and indicated how it was to be implemented. Chapters IV, V, VI, and VII dealt with the data, the construction of the discrete hierarchy, the estimation of the production function, and the estimation of the cost-share equations.

While the parameter estimates varied slightly when different variable sets were used, the acceptable hypotheses did not. Once the hierarchy was added to the model decreasing or constant returns to physical inputs prevailed. This contradicts the basic assumption of the models discussed in Part I of Chapter II. If the hierarchy variable represents phenomena external to the urban area, those models will have to be modified. The results suggest that recent literature, by ignoring factors external to the urban area (the hierarchy), has omitted a variable which could invalidate many of its results.

The finding of non-constant returns to scale has implications for neo-classical distribution theory. The basic postulate of urban models,
i.e., increasing returns to scale, invalidates the usual marginal product approach to distribution. Even though this study has found the source of increasing returns to be, in part, external to the urban area, the problem remains: paying factors their marginal products does not exhaust output. This study has attempted to determine which factors gain or lose. While this was done, the social or economic imperative that resolves the distribution problem was not explained. The resolution is an institutional phenomenon which economics has not analyzed in any depth. It is clear that some adjustments are being made, indeed, must be made; and, in the case of Saskatchewan, it appears that labour is the factor which gains most with urbanization.

In terms of the models discussed in Chapter II, this study is consistent with the approaches of Conrod (1972), Evans (1972), and Henderson (1974). The question as to which model of the urban hierarchy is best is unresolved, but it does appear that the stochastic approach can be rejected.

The policy implications of this study are not numerous. The study does, however, provide a firm base for interpreting and evaluating other work on Saskatchewan's urban system. Two questions which can be considered are the viability of individual communities and hierarchy levels and the basic policy strategy.

A reasonable objective or role for a government concerned with promoting population and economic growth is to initiate self-sustaining processes. If one had to select a town or hierarchy level on which to concentrate development policy, it would be preferable to select one which does not require constant injections of government aid, projects, or subsidies. The results of this study indicate that the communities in the fifth and third hierarchy levels (G5 and G3) would be the best places to concentrate development policy. As decreasing or constant returns always
prevail for G4, G2, and G1 it would not seem wise to attempt to promote
growth in those communities. While the paucity of observations for the
largest centres makes any comments dangerous, they appear to be at least
self-sustaining. If the results from the cost-share estimation are accepted,
and the data are reasonably reliable for the large centres, then the
communities of G76 appear to be growth centres and will grow as long as the
economy grows. 145

These results would seem to indicate that previous commentators have
been overly pessimistic about the growth prospects of communities with
populations of less than 100,000. 146 This study suggests that the critical
threshold level in Saskatchewan is between 5,000 and 10,000, the population
of the smallest cities in G76, and that the minimum size for continued
viability is about 1,000, the average population of G5. However, it must
be remembered that the production function estimation did not consider the
diseconomies of urban size, such as congestion, and ignored the issue of
returns to scale in urban infrastructure. Thus, our results will tend to
understate the minimum urban size necessary for viability or self-sustaining
growth.

145 In fact, it would appear that Regina and Saskatoon will continue to
grow even when the economy does not. Also, if the Prince Albert pulp mill
is any example, it would seem that single large investments are capable of
producing rapid growth in the communities of level G76.

146 Hansen (1970, 1972) argues that the optimum size city, for development
and growth purposes, is 250,000 to 750,000. In Hansen (1972) some authors
suggest 50,000 as the absolute minimum. In Robinson (1969) there is a
considerable variance of opinion. Some suggested 1,000 to 3,000 as the
minimum size for viability but, for sustained growth, the most commonly
suggested critical population was 100,000. This is a difficult problem
which could possibly be answered by extending the approach of the present
study.
With increasing returns for some of the hierarchy levels and constant returns for others, and perhaps decreasing returns for one level, it is likely that the hierarchy system in Saskatchewan will gradually be reduced to a three or four level hierarchy from its present seven levels. The point of this observation is that if a particular community is selected for development then the policy measures adopted must be such as to place the community into a viable, growing hierarchy level. If a community is not already in such a level then it will be necessary to embark on a thorough development project. It may be necessary to place an entire range of activities in a community, building a hospital and/or a school may not be enough.

Of course, the problem with raising a community into a growth hierarchy level is that such an attempt will, no doubt, dramatically alter the surrounding area's existing shopping patterns. The net effect may be to save Y at the cost of X and Z. The result of active intervention could easily be worse than doing nothing. The sensible approach is to concentrate on urban centres which can be identified as potential growth centres, to give each centre a push (if possible), and then let the natural economic forces of the system dictate the course of developments. A more active intervention requires a general equilibrium approach which integrates the production function, the urban hierarchy, shopping patterns, and transportation costs. But it is doubtful that even such a model could supply the information required to determine which communities should live or die. In the end, the long-run solution to small town Saskatchewan's continued existence lies in industrial expansion and population growth which can best be promoted by concentrating on existing growth centres, not by trying to create new ones.
Appendix I

Part I - Separability Constraints with Homogeneity

This appendix contains a derivation of the non-linear constraints imposed to test for separability. The linear constraints are straightforward, e.g., if capital and labour are separable from land then $\sigma_{KE} = \sigma_{LE} = 1$ which is equivalent to constraining $\gamma_{LE} = 0$ and $\gamma_{KE} = 0$ which implies that $\gamma_{EE} = 0$ and $\gamma_{KK} = -\gamma_{KL} = \gamma_{LL}$ when $\sum_j y_{ij} = 0$ for all $j$. The other linear constraints are similar. The non-linear separability tests require $\gamma_{ij} \neq 0$, $\gamma_{kj} \neq 0$ since $\sigma_{ij} = \sigma_{kj} \neq 1$. The acceptance of two types of separability requires the acceptance of the third since $\sigma_{12} = \sigma_{32}$ and $\sigma_{31} = \sigma_{21}$ implies $\sigma_{31} = \sigma_{32}$ since $\sigma_{ij} = \sigma_{ji}$.

The imposition of the non-linear constraints must be in terms of the remaining free parameters after the imposition of any other constraints. In Chapter VI the estimation is done in terms of $\alpha_L$, $\alpha_K$, $\gamma_{LL}$, $\gamma_{LE}$, and $\gamma_{KK}$. Remembering that $\gamma_{ij} = \gamma_{ji}$, the necessary and sufficient conditions for global separability, independent of the inputs, imply:

A. $\sigma_{KE} = \sigma_{LE}$ implies $\frac{\alpha_K}{\alpha_L} = \frac{\gamma_{KE}}{\gamma_{LE}} = \frac{\gamma_{KL}}{\gamma_{LL}} = \frac{\gamma_{KK}}{\gamma_{LK}}$

therefore

\[147 \text{ See Christensen and Berndt (1973a, pp. 86, 89-91) and (1973b).} \]
\[148 \text{ Ibid.} \]
\[149 \text{ Ibid., pp. 102-105 and Chapter III of this paper.} \]
\[ \sigma_K = \sigma_L \frac{\gamma_{KL}}{\gamma_{LL}} = -\sigma_L \frac{\gamma_{LE} + \gamma_{LL}}{\gamma_{LL}} \]

and

\[ \gamma_{KK} = \frac{\gamma_{LK}^2}{\gamma_{LL}} = \frac{\left[-(\gamma_{LE} + \gamma_{LL})\right]^2}{\gamma_{LL}} \]

B. \( \sigma_{KL} = \sigma_{EL} \) implies \( \frac{\sigma_K}{\sigma_E} = \frac{\gamma_{KL}}{\gamma_{EL}} = \frac{\gamma_{KE}}{\gamma_{EE}} = \frac{\gamma_{KK}}{\gamma_{KE}} \)

therefore

\[ \sigma_K = \sigma_E \frac{\gamma_{KL}}{\gamma_{EL}} \]

\[ \sigma_K = (1 - \sigma_K - \sigma_L) \frac{\gamma_{KL}}{\gamma_{EL}} \]

\[ \sigma_K = \frac{(1 - \sigma_L) \gamma_{KL}}{\gamma_{EL} + \gamma_{KL}} \]

\[ \sigma_K = \frac{(1 - \sigma_L)(\gamma_{LL} + \gamma_{LE})}{\gamma_{LL}} \]

and

\[ \gamma_{KK} = \frac{\gamma_{KE}^2}{\gamma_{EE}} = \frac{-(\gamma_{KK} + \gamma_{KL}) \gamma_{KL}}{\gamma_{EL}} \text{ since } \gamma_{EE} = \frac{\gamma_{EL} \gamma_{KE}}{\gamma_{KL}}. \]

\[ \gamma_{KK} \gamma_{EL} = (\gamma_{KK} - \gamma_{LL} - \gamma_{LE})(\gamma_{LL} + \gamma_{LE}) \]

\[ \gamma_{KK} = \frac{(\gamma_{LL} + \gamma_{LE})^2}{\gamma_{LL}} \]
C. \( \sigma_{LK} = \sigma_{EK} \) implies

\[
\frac{\alpha_L}{\alpha_E} = \frac{\gamma_{LK}}{\gamma_{EK}} = \frac{\gamma_{LE}}{\gamma_{EE}} = \frac{\gamma_{LL}}{\gamma_{EL}}
\]

therefore

\[
\alpha_L = (1 - \alpha_K - \alpha_L) \frac{\gamma_{LK}}{\gamma_{EK}}
\]

\[
\alpha_K = -\frac{\gamma_{EK}}{\gamma_{LK}} \alpha_L + \alpha_L - 1
\]

\[
= 1 - \alpha_L - \frac{\gamma_{EK}}{\gamma_{LK}} \alpha_L \quad \text{but} \quad -\gamma_{EK} = \gamma_{EL} + \gamma_{EE} = \gamma_{EL} + \frac{\gamma_{LE}}{\gamma_{LL}}
\]

\[
\alpha_K = 1 - \alpha_L - \frac{\gamma_{LE} + \gamma_{LE}^2 / \gamma_{LL}}{\gamma_{LL} + \gamma_{LE}} \alpha_L
\]

and

\[
\gamma_{KK} = -(\gamma_{EK} + \gamma_{LK})
\]

\[
= -\gamma_{EK} (1 + \frac{\gamma_{LL}}{\gamma_{LE}}) \quad \text{since} \quad \gamma_{LK} = \frac{\gamma_{EK} \gamma_{LL}}{\gamma_{LE}}
\]

\[
= (\gamma_{LE} + \frac{\gamma_{LE}^2}{\gamma_{LL}})(\frac{\gamma_{LL}}{\gamma_{LE}} + 1) \quad \text{since} \quad \gamma_{EE} = \gamma_{LE}^2 / \gamma_{LL}
\]

\[
\gamma_{KK} = \frac{(\gamma_{LL} + \gamma_{LE})^2}{\gamma_{LL}}
\]
Part II - Separability Constraints with a Non-homothetic Production Function

Separability of input $i$ and $j$ from $k$ implies

$$F_i F_{jk} - F_j F_{ij} = 0 \quad ^{150}$$

where $F_i$ and $F_{ij}$ are the first and second partials of the production function. For the translog this yields

$$\left( \frac{FM_i}{X_i} \right) \left( \frac{F}{X_j X_k} \right) (\gamma_{jk} + M_j M_k) - \left( \frac{FM_j}{X_j} \right) \left( \frac{F}{X_i X_k} \right) (\gamma_{ik} + M_i M_k) = 0$$

which implies

$$M_i \gamma_{jk} - M_j \gamma_{ik} = 0 \quad ^{151}$$

For LE-K this implies

$$\gamma_{EK} = \frac{\alpha_E \gamma_{LK}}{\alpha_L}$$

for non-linear separability

$$\gamma_{EE} = \frac{\gamma_{EL}}{\gamma_{LL}}$$

In the case of linear separability, $\gamma_{LK} = 0$ and $\gamma_{EK} = 0$. Similarly,

$^{150}$Berndt and Christensen (1973b).

$^{151}$Berndt and Christensen (1973a, pp. 102-103).
Linear  Non-Linear

\( LK-E \)
\[ \gamma_{EK} = 0 \]
\[ \gamma_{LE} = 0 \]

\[ \gamma_{KL} = 0 \]

\( KE-L \)
\[ \gamma_{LE} = 0 \]
\[ \gamma_{KL} = 0 \]

\[ \gamma_{KE} = \frac{\alpha_{K} \gamma_{LE}}{\alpha_{E}} \]
\[ \gamma_{KL} = \frac{\gamma_{KL}}{\gamma_{LL}} \]

\[ \gamma_{KK} = \frac{\gamma_{KE}}{\gamma_{EE}} \]
Appendix II

A POPULATION HIERARCHY VARIABLE

As was indicated in section III-A of Chapter VI, a population hierarchy variable (H2) should increase the chances of accepting decreasing returns to physical inputs and increase the relative importance of the hierarchy. This is indeed the case. In Table A1, columns (1) and (3) contain the estimates of the CD and Translog functions using L1 and H2; similar estimates for L2 and H2 are presented in columns (2) and (4). In all four cases decreasing returns to physical inputs are accepted at the 1% level while increasing returns to physical inputs and the hierarchy are accepted at the 1% level for all but the Translog using L1.\[152\] However, the Cobb-Douglas cannot be rejected at the 5% level, the computed $F(6,453)$'s were 1.84 and 1.98 for the L1 and L2 equations respectively.

\[152\] The t-statistics are:

<table>
<thead>
<tr>
<th>Physical Inputs</th>
<th>Physical Inputs and Hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>column (1)</td>
<td>5.61</td>
</tr>
<tr>
<td>column (2)</td>
<td>6.06</td>
</tr>
<tr>
<td>column (3)</td>
<td>5.49</td>
</tr>
<tr>
<td>column (4)</td>
<td>5.10</td>
</tr>
<tr>
<td></td>
<td>-4.43</td>
</tr>
<tr>
<td></td>
<td>-5.80</td>
</tr>
<tr>
<td></td>
<td>-0.764</td>
</tr>
<tr>
<td></td>
<td>-3.88</td>
</tr>
<tr>
<td></td>
<td>1. L1</td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
</tr>
<tr>
<td>α_L</td>
<td>.353727</td>
</tr>
<tr>
<td></td>
<td>(.0558009)</td>
</tr>
<tr>
<td>α_E</td>
<td>.0101259</td>
</tr>
<tr>
<td></td>
<td>(.259459)</td>
</tr>
<tr>
<td>α_K</td>
<td>.277422</td>
</tr>
<tr>
<td></td>
<td>(.0582427)</td>
</tr>
<tr>
<td>γ_LL</td>
<td>.329993</td>
</tr>
<tr>
<td></td>
<td>(.167593)</td>
</tr>
<tr>
<td>γ_EE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>γ_KK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>γ KE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>γ KL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>γ LE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>-.184879</td>
</tr>
<tr>
<td></td>
<td>(.0245188)</td>
</tr>
<tr>
<td>α_H2</td>
<td>.432583</td>
</tr>
<tr>
<td></td>
<td>(.0620115)</td>
</tr>
<tr>
<td>R^2</td>
<td>.9408</td>
</tr>
<tr>
<td>SSR</td>
<td>46.9964</td>
</tr>
<tr>
<td>DW</td>
<td>1.97</td>
</tr>
<tr>
<td>N</td>
<td>464</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


Mills, E. S., Urban Economics, Scott and Foresman, 1972, ch. 1 and ch. 7.


Urban Land Use in Ontario, Community Planning Board Ontario, Department of Municipal Affairs.
