Influence of Gross Regional and Industrial Product Ranks on Data Call Connections

Ian Geoffrey Kennedy

A thesis submitted to the Faculty of Engineering,
University of the Witwatersrand, Johannesburg,
in fulfilment of the requirements for the degree
of
Doctor of Philosophy

Johannesburg, 1992

DECLARATION

I DECLARE that this thesis is my own, unaided work. It is being submitted for the degree of Doctor of Philosophy to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other University.

I. G. Kenny

5 th day of October 1992

COPYRIGHT

Ian Geoffrey Kennedy

ABSTRACT

THIS STUDY identifies and evaluates factors that affect call connections in the South African public data networks, modelling these factors to aid data network planning. The research shows the relationship between the economic rank of each region served and the data communication resources required for that region. Moreover, it shows the resources required between regions.

THE THRUST of this thesis is that the volume of calls from a region can be estimated from its economic rank and more than 75% of the variation in the volume of calls between regions can be explained using the ranks of the originating and terminating regions. To prove this, records of more than four million calls are accumulated for all regions of the South African packet switched data network. An appropriate filtering and aggregation method is developed.

EXISTING growth models including the gravity model are separately examined. Based on probability and dimensional arguments, the Bell System growth model is selected. It is revealed that the success of this model depends on one premise being satisfied: this model tacitly and implicitly assumes that the originating and terminating calls are statistically independent.

RETURNING to the data network, it is found that the call connections (after filtering and aggregation) display dependence of destination on origin. Reasons for the dependence are discovered. Multiple linear regression reveals the nature of this dependence. Surprisingly, distance is not a factor. The importance of regional ranks and an inter-regional indicator variable are also discovered.

FINALLY, call volume from a node is shown to be directly linked with the weighted Gross Regional and Industrial Product of the region. This quantity, in turn, is inversely related to the rank of the region. Call connections are then modelled to be equal to the call connections within the first ranked region divided by the product of the originating region's rank and the terminating region's rank. This simple and economical model explains 76% of the variations that occur in call connections. It has proved its use by being included in the data transfer services product-line report.

ACKNOWLEDGEMENTS

THIS RESEARCH was financially supported by Telkom SA by way of a bursary, study time and office facilities.

I EXPRESS my grateful thanks to Doctor G. J. Rudolph for recommending the χ^2 test and to my mentor Professor Hugh Hanrahan for his helpful encouragement and support. Appreciation is also expressed to the anonymous referees, Professor Patrick Doreian of the University of Pittsburgh, Professor Robert Fildes of the University of Lancaster and Dr. Gordon Pirie of the University of the Witwatersrand for their valuable comments on and interest in portions of the thesis. Finally, thanks are due to Vic Read for proof reading.

DEDICATION

To Lee-Ann,

Cheral Dawn,

Gayle Julia and

Laura Kay

CONTENTS

Chapter 1
Background and Statement of Problem

Chapter 2
Previous Work in the Field

Chapter 3
Economic, Regional & Industrial Factors Affecting Calls

Chapter 4
Application of the Rank-size Rule

Chapter 5
Analysis of the Former Bell System Growth Model

Chapter 6
Role of Rank, Size and Distance

Chapter 7
Summary and Conclusions

LIST OF FIGURES

LIST OF TABLES

SUMMARY OF THEORY DEVELOPED WITH RESULTS

If 3 events A, B and C are statistically independent, then the joint probability of ABC is:

$P(ABC) = P(A)\ P(B)\ P(C)$.

In this thesis, $P(O_i)$, the probability of a call originating; $P(T_j)$, the probability of a call terminating and $P(D_{ij})$, the probability of a call traversing a distance D_{ij} are modelled as being statistically independent. That is, $P(ij)$, the probability of a call occuring between i and j is:

$P(ij) = P(O_i)\ P(T_j)\ P(D_{ij})$. $(R^2 = 0{,}847)$

From the definition of probability, $N_{ij}/S = [O_i/S]\ [T_j/S]\ P(D_{ij})$.

Simplifying, $N_{ij} = O_i\ T_j\ P(D_{ij})/S$.

(Thus, at a future point in time denoted by asterisks: $N_{ij}{}^* = O_i{}^* T_j{}^* P(D_{ij})/S^*$. By dividing the above two equations, the former Bell System growth model is proved — see chapter 5.)

Alternatively, a particular call connection, e.g. the number of calls within the busiest region, N_{11} can be calculated:

$N_{11} = O_1\ T_1\ P(D_{11})/S$. Otherwise the rank-size rule can be substituted:

If the rank-size rule applies — see chapter [illegible] — then $O_i = O_1/i$ and $T_j = T_1/j$. $(R^2 = 0{,}953)$

Thus, $N_{ij} = [O_1/i]\ [T_1/j]\ P(D_{ij})/S$.

Rearranging factors: $N_{ij} = O_1\ T_1\ P(D_{ij})/S\ /ij$

If $P(D_{ij}) = P(D_{11})$ — see chapter 6 — then

$N_{ij} = N_{11}/ij$. $(R^2 = 0{,}761)$

CHAPTER 1
BACKGROUND AND STATEMENT OF PROBLEM

1.1 BACKGROUND

MODERN SOCIETY relies on networks for communication, transport and to distribute energy and goods. The complexity and cost of these networks demand that existing networks be used effectively and that new networks be rationally designed. To meet this demand, the discipline of network planning has developed. This thesis is a contribution to the discipline.

Planning is the process of preparing a course of action to achieve goals. *Network planning* is the allocating of *resources* such as personnel and equipment in the most profitable way. Call volumes impact on both capital expenditure and revenue. First, with knowledge of the tariff, they allow revenue to be forecast, as Appelbe *et al.* (1988) note. Second, with knowledge of the daily traffic profile, call volumes allow loading to be determined. This thesis presents a simple, economical model for call volumes between regions of a network.

A network is composed of origins and destinations forming nodes which service regions. A *node* is the mathematical abstraction of a point at which a call starts or ends. These points are respectively, the *originating node*, i and the *terminating node*, j. Between each pair of nodes, there is an abstraction of the routing, here named a *link*. Fig. 1 shows a simplistic network of six nodes, (1) to (6), serving corresponding regions (1) to (6). Link (1)→(2) is an abstraction of the routing which carries all the calls from region (1) to region (2). The advantage of this abstraction is that the volume of calls on a link does not vary with routing rearrangements. Fig. 2 shows N_{ij}, the number of *call connections* from (i) to (j).

This chapter states the general problem of providing service cost-effectively and chooses data calls for study. It then gives the reader the background to the problem of planning a cost-effective data network. The chapter states the general shortcomings in data network planning and the specific need for a model of cause and effect. The chapter continues by posing the research problem to be solved and delimits the problem to be studied. In closing, this chapter gives an outline of the thesis and summarizes the chapter.

1.2 STATEMENT OF THE GENERAL PROBLEM

Policy decisions made by network providers have far-reaching consequences on the economy as a whole. In networks for transport or telecommunication, *trips* or *calls* are the indivisible units of traffic. An understanding of calls is crucial to the sound formulation of policy.

But what causes calls? On the one hand, what theoretical explanations are there? On the other hand, what empirical evidence can be found? Which external factors are important, and how strong is their influence? Who or what accounts for the creation of calls? Does the profile or spatial spread of these agents affect the number of calls?

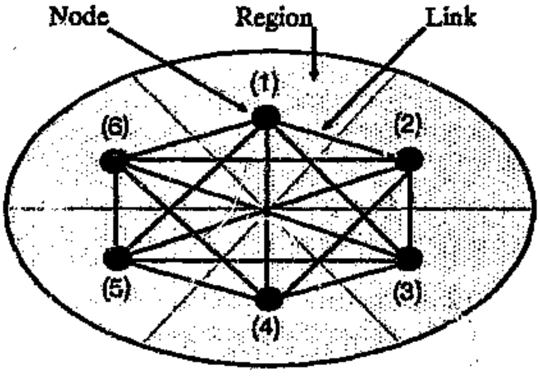

Figure 1. Regions, nodes and links

Specifically, what combination of factors and functions explains the origin and destination of calls?

1.3 CHOICE OF A NETWORK

Where can we gather evidence and how can we sift through it? Obtaining evidence requires that a network be selected for empirical investigation. All empirical researchers are faced with the apparent problem of obtaining solid data. They have to contend with random and systematic variations in measurements. They have to avoid inadequate sampling and missing data. A careful choice of study area can help here. The research must perforce be limited to networks where full call connection measurements can be made.

FROM \ TO	1	2	...	j	...	m	
1		N_{12}					O_1
2							O_2
:							:
i				N_{ij}			O_i
:							:
m							O_m
	T_1	T_2	...	T_j	...	T_m	S

Figure 2. Call connection table for figure 1

A closed network and a complete matrix is required for the research. This matrix must contain data on the major diagonal. Thus a national network was chosen for study.

Researchers in the transport field refer to fig. 2 as an *Origin-Destination (O-D) trip table.* A few recent doctoral theses in this field include those of Abu-Eisheh (1987), Alesifeer (1987), Almaani (1988), Chiu (1987), O'Neill (1987) and Rahi (1987). Networks for the transport of goods and people suffer from the drawback that it is impossible to fully capture how many trips are made from all origins to all destinations. To overcome this drawback, a telecommunication network was chosen that can provide data for all the cells in fig. 2. A *telecommunication network* is a network used to transport information, rather than goods or people.

Some modern telephone networks are able to sample origin-destination calls. However, these measurements cannot be made nation-wide or even region-wide. Nor can measurements be made simultaneously. If a network suitable for full instrumentation is chosen, then a large number of origin-destination counts can be collected by computer for subsequent analysis.

Data networks are telecommunication networks used to transport digital information between computers and/or terminals. Kennedy (1991b) reviews measurements available from the South African public telephone and data networks. The public data networks are good candidates for computer measurement as they have network management systems. But, as the function of each public data network is different, each network must be examined for suitability. For example, some networks offer *point-to-point* data communication; still others offer *point-to-service* data communication. The candidate networks for study were —

- Videotex network (Public Switched Telephone Network originating or X.28 originating);
- Analogue or digital dedicated data circuit network;
- Telex/teletex;
- Circuit-Switched Public Data Network;
- X.25 or Packet-Switched Public Data Network (PSPDN); and
- X.28 network (PSTN originating, terminating or both) using the PSPDN.

A short study of videotex total traffic was initially undertaken to obtain guidelines on which factors to select. Then attention was focussed on the PSPDN as the other networks do not have measurable call connections.

The particular network chosen for study is the public packet network in South Africa (X.25 plus X.28). (Data may be transferred between computers or computer terminals via the PSPDN. This X.25 network uses country-wide links efficiently by using each link in turn for many pairs of computers. This multiple usage is possible because the data from the originating computer is first divided into blocks of uniform size. An identifying label is added to each block, making a packet. Each packet is then switched to the destination specified in the label, where the blocks are put together to form the original data.) The X.28 service uses the PSTN to asynchronously access the X.25 network.

1.3.1 COST-EFFECTIVE DATA SERVICE PROVISION

According to Armstrong (1990), network engineering is the process of ensuring that a network meets the service requirements in a cost-effective way.

It is essential also that data network customers be served in the most cost-efficient manner. **Financial constraints** dictate that available capital be used effectively. Yet it is politic to provide a **satisfactory level of service** to all who want it. Earnings from existing equipment must be maximized by using the network to the fullest extent possible, without compromising the standard of service.

In this country, Telkom SA is currently the telecommunication administration that is the sole provider of telecommunication (non-broadcasting) service. The need to keep foreign exchange expenditure to a minimum leads to a re-examining of the methods used to plan and provide telecommunication services. The *network planning process* encompasses short-term (relief) planning and long-term (network expansion) planning of the regional allocation of resources. The delays and long *lead times* experienced, for example, in the acquisition of property and in the manufacture of telecommunication equipment mandate accurate allocation of the available resources.

Forecasts of call volumes can indicate the viability of new tariff schemes. A dearth of call data and a lack of models to predict calls may tempt the data network planner to try to be on the safe side, to over-dimension the network to ensure that no customer complaints ensue. However, the avoidance of customer complaints, the minimization of capital expenditure and the maximization of revenue can only be reached by continually monitoring the call patterns everywhere and keeping ahead of the demand. When *overloads* are detected, it might be too late to take corrective action because of the lead times involved.

Not all avenues of research have been explored since Jacobaeus (1980) sought research into "the nature of data traffic and of the dimensioning . . . of equipment". Unlike telephone networks, the problem of planning data communication networks is exacerbated by the high growth rates experienced. For example, during the year ending September 1990, the number of PSPDN ports provided to customers in this country increased by 15%, in spite of a generally quiet economy.

1.3.2 SHORTCOMINGS IN DATA NETWORK PLANNING

A planner of a data network is human and thus subject to the frailties of human nature. In his survey, Reason (1987) states this problem succinctly. The rest of this section is adapted from his excellent synopsis.

The planner of a data network has a short-term and a long-term *memory*. He or she keeps in short-term memory a data base of the country, the rapidly changing, complex network, plus methods that may be applied to the data. The working memory available for this data and relevant variables is limited.

The data network planner carries a similar structure of information and methods learned over a longer period. The planner bears a burden of *pet theories: out-of-date or incorrect data, and methods* that have not yet been "unlearned". Also, the planner's performance at recalling planning data and planning methods is imperfect.

The planner readily uses samples that are easily obtained from the data network, potentially causing *systematic biasing*. A planner has a deficient knowledge of important, relevant *statistical concepts* such as causality, correlation and confidence levels. He or she uses pet theories to estimate variables that are hard to obtain. The danger is that the planner loses sight of what was measured, what was estimated, and how good the estimating method is. Because the *feedback* process is over long periods, these problems may never be put right.

Finally, the *complexity* of planning a data network does not make for an easy job. Many opposing opinions and requirements have to be resolved and compromises found.

Perhaps the solution would be to supply the data network planner with suitable data bases and models. The next sections examine the problems found there.

1.3.3 LACK OF RELEVANT DATA

It needs only a small change of routing in the network to cause the number of calls at a point to change, all other things being constant. So, it is not sufficient to measure the number of calls at a point in a network. Anyway such bulk measurements give no hint as to where the calls come from or where they are going. The answer lies in using an abstraction of the network in which the physical routes are ignored, and the sum of calls from one region to another is considered.

Existing data bases have too much detail for the data network planner to be able to interpret what is happening. It is essential to exclude extraneous effects in the existing data.

Uncertainties beget over-provision. Call volumes had to be guessed by the data network planner. Consequently over-provision in the packet network was excessive. Although port provision is in multiples of the size of a packet switch, according to SAPT (1991), the number of PSPDN ports available in March 1990 was 9 909 when the demand was for only 3 399, a national over-provision factor of nearly three.

1.3.4 LACK OF RELEVANT MODELS

In engineering, the complexity of a problem may be so great that it is impossible to derive an exact solution. A *mathematical model* of the system which is a partial representation of reality, may afford adequately accurate answers. Unfortunately, the data network planner does not have the time or the talent to collect information, understand it, process it and produce models to aid the planning.

Stewart (1947) believes "The time to emphasize individual deviations is after the general averages have been established, not before." Finding the general averages requires data to be correctly collected and normalized. Normalizing is an important process that helps the planner see the individual peculiarities of the network.

Two classes of models exist. The first uses explanatory variables from **outside** the system, i.e., *exogenous* variables. The second class uses explanatory variables from **within** the system, i.e., *endogenous* variables. Wheelwright and Makridakis (1985: 38-41) define the two classes of models as being *time-series models* and *explanatory models.*

This raises a very interesting philosophical point. If we have a good explanatory model, we do not need a time-series model, because we can predict future values by inserting the updated parameters in the explanatory equation! The problem with time-series modelling is that it is a bit like trying to predict all the share prices on the stock exchange without reading and researching what is happening in the real world. The approach taken in this thesis is to consider the universe in which the data is immersed, establishing the factors that are at work in [illegible]conomy driving the data. Of course, the challenge then becomes to find what relationships exist.

Using the explanatory class of model, Defris *et al.* (1986) show that telecommunication traffic follows aggregate economic activity, in particular, leading indicators. In the field of data communication, Terranova *et al.* (1989) show that the logarithm of the number of data terminals in a country is linearly related to the logarithm of the aggregate economic activity in the country. Given sufficient historical data, exogenous variables might be used to model the volume of data calls in the country. Models using exogenous variables require historical records of the volume of calls between all regions to have been kept over time. But the data network planner needs a method now.

The second class of model, using endogenous variables, is useful in the absence of historical data. The *CCITT* is the highest international body of official advice to telecommunication administrations. According to the CCITT (1983), the absence of traffic measurements is unfortunately quite common in developing countries. Although the CCITT is referring to telephone traffic, its model may be applied to data calls. In the absence of traffic measurements, the CCITT (1988) advises that the *equidistribution model* be used. This model is originally from Rapp (1962). The model assumes a *community of interest factor* between each pair of regions. That is, no attempt is made to explain the traffic that exists among regions.

1.3.5 ABSENCE OF CAUSE-AND-EFFECT MODELS

There are sources of exogenous and endogenous variables from which a data network planner could draw individual figures. Unfortunately, planners are unable to assess the relative importance of these variables and model how they affect data calls. Mathematical models for data call volumes within this country are conspicuous by their absence, and there is a suspicious silence in the international literature. An explanatory model would help the data network planner understand the forces at work in the network. A *cause-and-effect* or *explanatory* or *causal model* treats the system being studied as a "black box" with an output ("effect") that depends upon the inputs ("causes"). The planner can find out how the output is sensitive to variations in the input. Such a model also lets a planner see the role and relative importance of variables. Such a model would explain in general terms the underlying reasons for the observed phenomena. Most importantly, the model can be used to gain a better understanding of the workings of the system.

(One application of an explanatory model is for the data network planner to use it as a normative model. A *Normative model* is a model that is used as a norm against which measured values can be compared and reasons sought for the discrepancies.)

1.4 STATEMENT OF THE SPECIFIC PROBLEM

In packet switched networks, direct self-sufficient routes are not provided. This means that, even if they were analysed, historical bulk measurements would be of no worth when regional boundaries are changed or new centres established. It is also not obvious how call volumes change as the area served by a node grows unevenly or evenly. Although raw data is available as a by-product of the call billing process, no analysis of call connections for either the X.25 or X.28 services had been undertaken.

To be able to plan equipment quantities and its deployment, the planner of a network must know the current call connections between nodes and be able to predict what they will be in the future. Planners of networks must have all call connections (all cells in fig. 2) to do their planning. If the calls measured from (1) to (2) and from (1) to (3) in fig. 2 are carried on the same route, it is easy for the planner to add N_{12} and N_{13} to determine the *bulk traffic*. The converse is not true without a model.

Even if the data network planner was aware of models derived for other networks, contradictions between the models would cause the planner to mistrust the application of the models to data networks. Models from other disciplines demand a sound theoretical basis and require validation before being acceptable to a data network planner. The research problem is therefore **to construct a theoretically valid and empirically accurate model for call connections** to place the planning of data networks on a sound foundation

1.4.1 DELIMITATIONS

The research does not cover the *performance modelling* of the data networks. Instead it seeks to provide an explanation for the calls observed. The output of the model then gives the inputs needed for a network performance model.

The study includes all national calls. This enables a *closed system* to be studied where the number of calls into the network is equal to the number of calls out of the network.

1.4.2 STATEMENT OF GENERAL APPROACH

The procedure will be to find suitable candidates for external variables; within a limited time-scale, measure call volumes in the South African public packet network and determine the relationships. The information obtained and the methods developed should aid the rational design of data networks and the extension of existing networks.

The approach proposed above is not without its risks. First, suitable factors may be overlooked. Second, external or network data may prove unavailable or be too unreliable for use. Third, unpredictable variations in the network data may render the data useless. Fourth, a matching of external and network regions may prove impossible. Fifth, it is possible that no meaningful relationships may be discernible in the data.

1.5 HYPOTHESES

The quest is to find interlocking hypotheses from the specific problem. Six are postulated:

The first hypothesis: Average regional call volumes are proportional to weighted economic activity in the region. The idea of introducing weighting is novel.

The second hypothesis: Weighted and unweighted economic activity in this country follow the rank-size rule. This hypothesis has to be tested for the country being studied.

The third hypothesis: Call connections are proportional to both originating and terminating call connection totals. This is not obvious -- chapter 2 reveals contradictory models.

The fourth hypothesis: Distance between originating and terminating regions must be invoked as an exogenous variable to explain call connections. Chapter 2 suggests otherwise.

The fifth hypothesis: Call connections decrease with both the rank of the originating region and the terminating region. This hypothesis introduces a new way of looking at old data.

The sixth hypothesis: Calls connection volumes between regions differ from those within a region. This hypothesis is a very subtle point, and could be the basis for further research.

1.6 PLAN OF THE STUDY

The study is divided into seven chapters. Chapter one introduces the study. The remaining chapters cover the conduct and course of the research and the conclusions:

Find suitable models: Similar problems to the current problem exist in the fields of telegraphy, telephony, transport and geography. Chapter two reviews related research models that might be used to explain call volumes in the South African data networks. It also contains a review of techniques basic to this thesis.

Find suitable variables: Chapter three finds sources of exogenous variables to explain call volumes in the South African data networks.

Generalize the variables: Chapter four generalizes the exogenous variables to any network.

Obtain call data: Chapter five finds what raw data is available for the South African data networks and develops techniques to make measurements of call connections available. It compares the measurements with the most promising model from chapter two.

Find the strength of the factors: Chapter six explores which variables are important. It finds the strength and direction of each factor.

Develop a model: Chapter seven brings together the ideas presented and includes ideas for future research.

1.7 SUMMARY

It is now appropriate to summarize what has been found about the problem.

Financial constraints demand that data network planning be placed on a sound scientific basis. Resources purchased with scarce capital must be deployed where the equipment is needed most. The scientific method of deciding where equipment must be deployed is to make measurements of current call patterns and to develop a theory of cause and effect.

Most specifically the problem is to find laws, as simple and as general as possible, that explain the data and to create models that adequately describe the observed calls. The human nature of planners, their limited memory, imperfect methods and data, and the complexity of the task make it imperative that the models be easily understood.

In this chapter, the background to the research has been discussed first. The statement of the research problem and the hypotheses have been formulated. The delimitations of the study were given and finally the plan of the thesis has been presented.

A full review of the literature follows in the next chapter.

CHAPTER 2
PREVIOUS WORK IN THE FIELD

2.1 INTRODUCTION TO THE LITERATURE REVIEW

MANY METHODS of modelling call connections may be found. These are the matrix method, the interest techniques, the gravity models, the growth models, the ideal call distribution method, the double-factor procedure and the time-series models. These models differ in the number and type of influences assumed to be at work in creating call connections.

The matrix method says in effect: "Here are the regions; this is the number of calls between each." The method makes no attempt to explain the number of call connections. "All we can do is measure or guess the numbers and put them in the table."

The interest techniques say: "Let's normalize the number of calls on a link. We hope that each will then be constant." In particular, the number of calls may be singly or doubly normalized.

The gravity models assume that calls are not caused solely by the sizes of the centres sourcing and sinking the calls, but also by the space, D_{ij} separating them.

Growth models assume that the increase in call connections can be modelled only in terms of the size of the origin and destination and the growth of each.

The ideal call distribution method presumes that calls are dispersed perfectly in proportion to the size of the origin and the destination of the calls.

Kruithof's double factor method uses iterative scaling to form a growth model that projects an ideal call distribution matrix into a future matrix.

Time-series models assume all call connections have been measured over a period of time and extrapolates the observed sequences.

This chapter critically reviews literature relevant to the study. The rest of this chapter is divided into two parts. The first deals with methods, techniques and models from the different scientific disciplines. The second focuses on the specific study area (telecommunication).The chapter concludes by choosing for further study the most promising model from the literature review.

2.2 MATRIX METHODS

Newstead (1961) uses a 2-dimensional *matrix* as a repository for the storage of point-to-point call counts or traffic. Though in practice, this structure must be enlarged to include O_i and T_j, the originating and terminating totals, and their place names.The number of call connections can be entered into a *cell* of this table. Two criteria or *factors* decide how the measured call connections are to be classified: origin and destination. Therefore the table containing call connection counts is a *two-way table*, here a m^2 table, where m is the *number of regions*. Scheaffer and McClave (1982) explain that a common question is, "Does the row criterion depend on the column criterion?" In other words, "Is the row criterion contingent upon the column criterion?" Then such a table is known as a *contingency table*. Chapter five takes up this matter further as it decides how origin and destination are dependent.

Allison (1985) feels that a *compromise of two matrices*, one for the morning and one for the evening of the busy season are adequate for most requirements. The steps involved in using the matrix for modelling and prediction are:

(1) create a base matrix,

(2) grow the matrix,

(3) compare top-down and bottom-up forecasts.

BASIC NOMENCLATURE USED IN THIS THESIS

(a) Scalar quantities (see fig. 2)

S = total number of calls in the network (also the sample size in this thesis).

i = subscript of originating region (redefined as rank of originating region).

j = subscript of terminating region (redefined as rank of terminating region).

(b) Vector quantities (see fig 2.)

O = vector containing the number of originating calls.

T = vector containing the number of terminating calls.

(c) Matrices

N = matrix of the number of call connections between regions at the time of the study.

N^* = matrix of the no. of call connections between regions at a future point in time.

G = matrix containing the growth ratios, defined such that $G_{ij} = N_{ij}^*/N_{ij}$.

U = matrix containing the usage, defined such that $U_{ij} = (N_{ij} + N_{ji})/2$.

P = matrix containing the probability of a call connection occurring between regions.

F = matrix containing the community interest co-efficients N_{ij}/S.

K = matrix containing the co-efficient of interest, such that $K_{ij} = SN_{ij}/(O_iT_j)$.

D = matrix containing the distances between the major metropolitan centres.

2.3 INTEREST TECHNIQUES

In this thesis, the sum of the elements of the matrix will be represented by the symbol S. This will also be the *sample size*. S is the *total number of calls* flowing in the network. See fig. 2. Wood *et al.* (1974) advise "**Non-dimensional** parameters have the great advantage of generality". That is, they transcend local trends and parochiality such as local units of measurements. Taking this argument further, Saaty and Alexander (1981) point out that if a person divides a matrix by S, it renders the matrix unique. This could be a useful technique, giving the *call probability* P(ij) rather than the number of call connections, N_{ij}. However, this thesis will use the call connection, paying deference to the engineer's preference for dimensioned variables. Liang (1988) takes an entirely different attitude.He defines *community interest coefficient of i and j* as $F_{ij} = N_{ij} / S$. Each F_{ij} is assumed constant with time. This is tantamount to assuming that P(ij), the probability of a call connection occurring, is static. Liang's method is equivalent to a single normalization of the call connections. Double normalization will be discussed below under the heading of telecommunication techniques.

2.4 THE GRAVITY MODEL

Transport engineers are vitally interested in what they term *Origin-Destination (O-D)* tables. Drew and Chen (1976) point out various areas where transport planning can be applied: vehicular O-D trips, harbours, highway growth, public transport, vertical access in tall buildings and urban development. The methods used by transport engineers are invariably extensions of the gravity model and are examined next.

The naive gravity model is deeply ingrained in literature.It has been used to model a variety of events, ranging from calls via telecommunication and transportation networks to migration and marriage distance. However, it suffers from intrinsic flaws: It is a misapplication of a 3-dimensional equation to a 2-dimensional problem and suffers from vagueness in the definition of distance. Many modified gravity models are dimensionally unsound. This section shows places in literature where disquiet is declared.

Telecommunication engineers ask: How many calls do people make from one area to another? Sociologists are interested in the flow of people or the flow of their goods between one region and another. Sociologists wonder: Why does this number of people move from here to there? Similarly, workers in the field of transport ask: Why do people send this number of goods from one region to another? Can the number of calls or trips or number of goods be modelled? Engineers, sociologists, geographers, and geneticists share an interest in networks and use mathematics as a tool to model these flows.

Leik and Meeker (1975) believe that sociologists and others have been attracted to mathematics because of its *precision* and its power to *abstract*. For example, the precision of an equation enables tests to be performed to establish whether experimental data conform with a model. Abstraction of data into an equation enables factors and influences to be stated succinctly and unambiguously.

The potential applications for mathematical modelling of flows through networks are legion. In the telecommunication field, Jipp (1960) and Dunstan (1977) model telephone calls, with Palmowska (1981) modelling telex calls. Taylor (1971) refers to studies of the spreading of viruses, the dispersion of genes, the distance between a person's birthplace and place of marriage, the migration of people or their journeys to work. While Whitmore (1965) reports on urban driving, Byler and O'Sullivan (1974) model the movements of trucks. Furthermore, Fotheringham (1981) models air-trips, leisure scientist S. L. J. Smith (1985) models recreational trips, and sociologist T. S. Smith (1976) concentrates on a criminal's travel between crime-scene and residence. Knox (1980) models the accessibility of primary health care and Allen (1988) models historic patterns of trade. So, wherever people move or convey goods, services and information, the gravity model has been applied.

This chapter shows how dimensional analysis and probability analysis lead to a theoretically sound model.

In the next section, the basic gravity model and some of its variants are discussed. The following section examines a different approach. The final section argues against gravity models.

2.4.1 EXTENSIONS TO THE GRAVITY MODEL

A model based on the equation from Newton's Law of universal gravitation is termed a *gravity model*. However it is just a model, not a law: the assumptions underlying the model must have a theoretical basis. "Analogy may generate insight, but it does not contribute toward validation" cites Tocalis (1978). It is easy to apply Newton's equation blindly to other fields of science. It is harder to use Newton's method of observation, testing, and rationalization.

The basic model: The textbook gravity model of Oppenheim (1980) assumes (without proving it) that N_{ij}, the number of goods or calls transmitted from region i to region j in the study period, is proportional to the size of the region i and to the size of the region j. It also assumes that the transfer is inversely proportional to the square of the *distance* D_{ij} separating the regions. That is,

$$N_{ij} = k\, O_i\, T_j \,/\, D_{ij}^2. \qquad (1)$$

Here k is a constant of proportionality. O_i and T_j are the *sizes of the originating and terminating regions*.

However, Curry (1972) points out that a "vague disquiet" prevails in literature about gravity models. This is worthy of further investigation.

Mathur and Satsangi (1985) cite criticism of equation 1. They feel that "even the basic hypothesis of the model is unreasonable... The model is logically inconsistent and does not give a good explanation of observed traffic flows". They believe that there may be a need to use "different techniques, (and) from a basis of logically sound hypotheses, to develop models...(that)...explain the underlying behavioural phenomena satisfactorily".

It seems that the basic gravity model cannot be expected to fit all networks well. Byler and O'Sullivan (1974) use equation 1, studying only trips to one destination. They conclude that "the chief inaccuracy lies in the tendency to underestimate major flows." On the other hand, Christian and Braden (1966), using the same equation find that it underestimates for minor destinations.

Jipp's slight modification: The basic gravity model has spawned many modifications. A *modified gravity model* is one in which D_{ij} still appears. If the basic model does not fit, then the single most obvious extension is to modify the exponent. This gives:

$$N_{ij} = k\, O_i\, T_j / D_{ij}^{c}. \tag{2}$$

Jipp (1960), for example, advocates a slightly modified gravity model for the telephone traffic volume between European countries. He adopts a distance power, c, of 2½. Subsequently, Davis (1965) gives four further examples of this equation. These are gravitational or electrostatic forces, inter-molecular forces, intra-molecular forces and forces within the nucleus. This means that nature does not prefer inverse square laws. His exponents are approximately 2, 7, 10 or 15.

More recently, S. L. J. Smith (1985) believes that although some researchers characterize equation 1 as naïve, it has some merit. He proceeds to use equation 2 and suggests that the resulting "model misspecification is not a serious problem"

Most recently, Mueser (1989) uses equation 2, with c = 1. He concludes that the "deviations from the...model appear stable over time (and this) is particularly troublesome".

Fotheringham's further modifications: Fotheringham (1981) decides to work elsewhere. He raises Tj to a power g_i that depends on i, and raises D_{ij} to a power c_i that also depends on i. That is,

$$N_{ij} = k\, O_i\, T_j^{g_i} / D_{ij}^{c_i}. \tag{3}$$

He concludes that "empirical research suggests that estimated distance-decay parameters are a function of spatial structure". He defines spatial structure as meaning "the size and configuration of origins and destinations" This circular argument shows us that his further modifications to the gravity models are a failure. Later, Fotheringham (1984) admits that "gravity interaction models (like equations 1 to 6 above) are misspecifications of reality because they fail to model accurately". He uses empirical evidence to show the resulting bias and advocates that spatial structure be used as a further explanatory variable.

Congdon continues modifying: Congdon (1989) goes a step further. He raises O_i to a constant power f and Tj to a constant power g, giving:

$$N_{ij} = k\, O_i^{f}\, T_j^{g} / D_{ij}^{c}. \tag{4}$$

Isard's asymmetric approach: Hua *et al.* (1979) cite Isard's approach to further generalizing equation 4. Isard allows for weighting parameters, u_i and v_j:

$$N_{ij} = k\, u_i\, O_i^{f}\, v_j\, T_j^{g} / D_{ij}^{c}. \tag{5}$$

Smith's subtraction: Sociologist T. S. Smith (1976) takes a different path. His model giving the best fit subtracts a constant, N, from the quotient:

$$N_{ij} = (k\, O_i\, T_j / D_{ij}) - N. \qquad (6)$$

This illustrates the futility of using regression analysis on models that have no theoretical foundation.

Anderson's addition: Anderson, reported in Rich (1980), adds a constant, d, to D_{ij} to avoid problems on the major diagonal where D_{ii} is zero. He models:

$$N_{ij} = k\, O_i\, T_j / (D_{ij} + d). \qquad (7)$$

This avoids the difficulty of dividing by zero when dealing with calls from a region to the same region. But it comes dangerously close to distorting the definition of distance.

Dunstan's difficult equations: Dunstan (1977) takes the gravity model the furthest from its original form. He finds that the distance exponent is a function of r_i, the telephone service density measured in region i. He introduces a coefficient of interest, K_{ij}, to bring observed and predicted values into line. This yields

$$K_{ij} = k_{ij}\, a\, b^{W} (c/D_{ij})^{\{1/(d r_j)\}^{e}}, \qquad (8)$$

$$\text{with } N_{ij} = (O_i T_j / S)\, K_{ij}\,. \qquad (9)$$

Here S is the total network traffic, and a, b, c, d, e and each k_{ij} is a constant. W is unity if regions i and j are adjacent and if j is away from the city centre and if j has less than 2 000 services. W is zero otherwise. Finally, r_i equals the number of services per hectare in region i.

Although Dunstan's equations commendably use dimensionless parameters, his model is too specific to apply directly to any other network.

2.4.2 THEORETICAL BASIS OF GRAVITY MODELS

Equations 1 to 8 are deductive models, not theoretical laws supported by empirical evidence. No theoretical basis has been advanced for the adoption of these equations. On the contrary, six arguments can be used against them:

Misapplication arguments The least compelling argument relates to misapplication. Newton's law arises from the 3-dimensional integral of a potential that drops off inversely with distance. See for example Purcell (1965). Newton's law applies only to spheres whose mass is distributed spherically symmetrically. See for example Kittal *et al.* (1965). The gravity model

$$N_{ij} = (k\, O_i\, T_j / D_{ij}) - N. \tag{6}$$

This illustrates the futility of using regression analysis on models that have no theoretical foundation.

Anderson's addition: Anderson, reported in Rich (1980), adds a constant, d, to D_{ij} to avoid problems on the major diagonal where D_{ii} is zero. He models:

$$N_{ij} = k\, O_i\, T_j / (D_{ij} + d). \tag{7}$$

This avoids the difficulty of dividing by zero when dealing with calls from a region to the same region. But it comes dangerously close to distorting the definition of distance.

Dunstan's difficult equations: Dunstan (1977) takes the gravity model the furthest from its original form. He finds that the distance exponent is a function of r_i, the telephone service density measured in region i. He introduces a coefficient of interest, K_{ij}, to bring observed and predicted values into line. This yields

$$K_{ij} = k_{ij}\, a\, b^{W} (c/D_{ij})^{\{1/(d r_i)\}^{e}}, \tag{8}$$

$$\text{with } N_{ij} = (O_i T_j / S)\, K_{ij}\,. \tag{9}$$

Here S is the total network traffic, and a, b, c, d, e and each k_{ij} is a constant. W is unity if regions i and j are adjacent and if j is away from the city centre and if j has less than 2 000 services. W is zero otherwise. Finally, r_i equals the number of services per hectare in region i.

Although Dunstan's equations commendably use dimensionless parameters, his model is too specific to apply directly to any other network.

2.4.2 THEORETICAL BASIS OF GRAVITY MODELS

Equations 1 to 8 are deductive models, not theoretical laws supported by empirical evidence. No theoretical basis has been advanced for the adoption of these equations. On the contrary, six arguments can be used against them:

Misapplication arguments The least compelling argument relates to misapplication. Newton's law arises from the 3-dimensional integral of a potential that drops off inversely with distance. See for example Purcell (1965). Newton's law applies only to spheres whose mass is distributed spherically symmetrically. See for example Kittal *et al.* (1965). The gravity model is thus a misapplication to 2-dimensional space. Also, the origins of the calls are not spherically symmetrically distributed in a sphere.

Dimensional analysis in modelling It is necessary to remind the reader of the importance of dimensional analysis as a tool when modelling. No equation used in modelling may be dimensionally invalid. Housner and Hudson (1959) for example, cover *dimensional analysis*, the analysis of units of measurement, used especially in physical modelling. In modelling, both sides of every equation must be expressed in the same primary units. The authors state that suitable systems of primary units include the Length-Mass-Time or L-M-T system (used in physics) and the Length-Force-Time or L-F-T system of units (used in engineering). However, here in the field of network modelling, an appropriate system of primary units would be Length-Calls-Time or L-C-T. In dimensional analysis, a physical quantity can only be related to other physical quantities in certain specific ways. For example, see Neal and Shone (1976). Derived units must have the form of products of integer powers of the primary units. According to this analysis, equations 2 to 5 cannot be dimensionally correct if the power of D_{ij}, O_i or T_j is not an integer.

Calls within region, N_{ii} The gravity model fails to correctly model the number of calls within a region. As the $D_{ij} \to 0$, $N_{ii} \to \infty$, an impossible and unrealistic value.

Calls symmetric about the major diagonal, N_{ij} and N_{ji} As $D_{ij} = D_{ji}$ and $O_iT_j = T_jO_i$, the gravity model predicts that $N_{ij} = N_{ji}$. This is overly simplistic.

The arbitrariness of D_{ij} The distance D_{ij} is presumably the distance between the centre of region i and the centre of region j. But how do researchers choose the centre of a diffuse region? They may base it on the co-ordinates of the major city associated with the region or even adjust D_{ij} to make the gravity model fit better. For example, see equation 3.

Probability argument against the gravity model Equation 10 assumes that P(ij), the probability of a call being on a link i-j is equal to the product of the probability P(i) of a call originating in one region and the probability P(j) of a call terminating in another region, multiplied by S, the sum of all calls in the network:

$$P(ij) = P(i).P(j).S \tag{10}$$

The equation is deliberately defective dimensionally. The left hand side is dimensionless, while S — and hence the right hand side — is measured (e.g.) in units of calls per month. By definition of probability, equation 10 is the same as:

$$N_{ij}/S = (O_i/S).(T_j/S).S \tag{11}$$

Simplifying,

$$N_{ij} = O_iT_j. \tag{12}$$

Therefore at the future point in time, indicated by an asterisk,

$$N_{ij}^* = O_i^*T_j^*. \tag{13}$$

Dividing, equation 13 by equation 12 and using the definitions of the growth ratio gives the growth ratio $G = N^*/N$:

$$G_{ij} = G_iG_j. \tag{14}$$

Equation 14 is defective because it is derived directly from the dimensionally defective equation 10.The units on the left and right side of equation 14 are different.

Models such as the modified gravity model of Jipp (1960) assume an equation of the form:

$$N_{ij} = k\ O_iT_j/D_{ij}^c, \tag{15}$$

Here k and c are constants, and D_{ij} is the distance between origin and destination. Thus the number of calls at a future point in time is:

$$N_{ij}^* = k\ O_i^*T_j^*/D_{ij}^c. \tag{16}$$

Dividing equation 16 by equation 15 also gives equation 14. But, equation 14 is flawed because it was derived from equation 10. Therefore the classical gravity model of equation 15 is fatally flawed.

2.4.3 GRAVITY MODEL CONCLUSION

The ubiquitous gravity model in the literature suffers from fatal flaws: It is a misapplication of a 3-dimensional equation to a 2-dimensional problem. The modified gravity models are dimensionally unsound or suffer from vagueness in the definition of distance. The use of probability and dimensional analysis shows that the gravity model must be abandoned.

2.5 GROWTH MODELS

Let G_{ij} represent the growth ratio, N_{ij}^*/N_{ij}. Let G_i represent the growth ratio O_i^*/O_i, and let G_j represent the growth ratio T_j^*/T_j. Here the asterisk indicates the value at a future point in time.

Without revealing the probability assumptions implicit in it, Newstead (1961) mentions the former Bell System growth model:

$$G_{ij} = G_i \, G_j / G. \tag{17}$$

However, he reveals that the standard forecasting method used in Australia is based on another model:

$$G_{ij} = (O_i \, G_j + T_j \, G_i) / (O_i + T_j). \tag{18}$$

Equation 18 is a mixed causal and growth model which has no probability foundation.

David and Pack of the then Bell Telephone Laboratories (1979) also suggest equation 17, again without stating the probability implications of it. They list other possible growth formulae:

$$G_{ij} = \tfrac{1}{2}(G_i + G_j); \tag{19}$$

$$G_{ij} = G_i \, (G_j)^{1/2}; \tag{20}$$

$$G_{ij} = (G_i \, G_j)^{1/2}; \tag{21}$$

$$G_{ij} = G_i \text{ or} \tag{22}$$

$$G_{ij} = G_j \,. \tag{23}$$

Equations 19 to 23 can be seen to be all members of the class $aG_i^m + bG_j^n$, where a, b, m and n are empirical constants. These equations will now be considered in the reverse order to that above.

Consider now a limiting case. Assume that the originating region, i, dies away. All the calls from that region to another region j must also die away. That is, in the limiting case,

$$\text{as } P(i) \rightarrow 0, \text{ so must } P(ij) \rightarrow 0. \tag{24}$$

This implies that P(ij) must be some function of P(i) and not just a function of P(j). So an equation such as 25 is incorrect:

$$P(ij) = P(j). \tag{25}$$

From the definition of probability,

$$N_{ij}/S = T_j/S, \text{ or} \tag{26}$$

$$N_{ij} = T_j \,. \tag{27}$$

Equation 27 is incorrect because it assumes that calls to destination j only come from i. At the future point in time, equation 27 implies equation 28:

$$N_{ij}^* = N_j^*. \qquad (28)$$

Dividing equation 28 by equation 27, gives equation 23, which then must be incorrect, because the above analysis shows it is based on erroneous equations 25 to 28.

A similar, symmetrical analysis will show that equation 22 is faulty.

Assume instead that equation (29) is correct:

$$P(ij) = \sqrt{P(i)P(j)}. \qquad (29)$$

There are no probability arguments to support equation 29. By definition of probability,

$$P(ij) = N_{ij}/S = [(N_i/S)(N_j/S)]^{1/2}. \qquad (30)$$

$$\text{Simplifying, } N_{ij} = [N_iN_j]^{1/2}. \qquad (31)$$

Therefore at the future point in time, $N_{ij}^* = [N_i^*N_j^*]^{1/2}$. (32)

Dividing equation 32 by equation 31 gives equation 21. Thus equation 19 rests on an unsubstantiated probability statement, equation 29.

No probability statement could be found for equation 20 or equation 19. These two equations are also suspect.

Summarizing, probability and dimensional analysis suggest that the most promising equation is equation 17. There is however an independence assumption underpinning this equation. The assumption was investigated empirically and is reported in chapter five.

2.6 TELECOMMUNICATION TECHNIQUES

In this section, techniques used in telephone and telex networks will be analysed. These may be regarded as attempts to break away from the gravity model and avoid determining the function of D_{ij}.

2.6.1 IDEAL CALL DISPERSION MODEL

Newstead (1961) defines the concept of *ideal call dispersion*. If the interest pattern among customers' terminals and computers is uniform over the network, the number of calls from a region will be distributed among all regions in proportion to the terminating size of each, T_j. This he terms the *ideal distribution*. He then introduces a community of interest factor which is the ratio of the observed to the ideal traffic. Community of interest factors are assumed to remain constant.

The CCITT (1983) defines the idea of a *community of interest factor*. This is a repetition of the community of interest factor of Rapp (1962): Rapp's approach is that if the number of call connections between two centres is unknown, it must be estimated by analogy with the number of call connections that were measured between another similar pair of centres.

Bear (1976) documents a typical approach in telecommunication to modelling the distribution of calls. The approach assumes a simplified law of the general form:

$$N_{ij} = O_i \, T_j \, K_{ij}, \tag{33}$$

where K_{ij} is a factor measuring so-called community of interest. His *community of interest factor* is assumed to remain constant over a prediction period. It glosses over the separate contributions of distance and true interest. The drawback of this approach is that the number of K_{ij} coefficients to be tracked increases as the square of the number of regions. If historical records are not available, predicting K_{ij} becomes impossible.

Palmowska (1981) also adopts a general model of the form given by equation 33. He regards O_i and T_j as parameters that quantitatively describe the originating node, i and the terminating node, j. The interest measure is a function of the distance D_{ij} and parameters that we hope are constant. He multiplies by a factor to adjust for non-distance specific behaviour for each ij pair. The whole Polish *telex* network is covered, but internal traffic is excluded. He takes O_i or T_j proportional to the users or better still, the traffic. It is one of the few studies made on a country-wide data-carrying network. His paper confirms that the number of telex calls decrease with distance traversed and he favours an asymmetric exponential model to a gravity model.

2.6.2 KRUITHOF'S DOUBLE FACTOR PROCEDURE

Kruithof (1937) is remembered for his pioneering work in determining a theoretical or *ideal call distribution* in a network. Kruithof's method is equivalent to a time-series model for only two points in time: now (known data values) and future (data values to be determined). He realizes that a call connection does not depend only upon the size of the originating exchange, but on both the originating and the terminating exchange sizes. Using both, he distributes the traffic into the matrix in proportion to the sizes of both the originating traffic and the terminating traffic, which may be different. Although Kruithof is remembered for this *double-factor method*, he also addresses matters such as hourly and seasonal *variations*.

Furness (1965) describes a technique "first proposed at the IBM seminar in Utrecht". The technique described is Kruithof's method. Furness contributes nothing new. Thus the method is best known as Kruithof's method.

Debbisse and Matignon (1979) compare different ways of *matrix projection*. To estimate unknown call connections, they advocate Kruithof's method, as do Combot *et al.* (1988). Kruithof's double factor method is merely one way of filling each cell of the matrix with a determinate but nevertheless arbitrary number. His iterative method has a much higher overhead than equation 17, and has no theoretical basis.

2.6.3 TIME-SERIES MODELS[1]

The CCITT (1968:12) gives the *logistic* and *Gompertz* saturating functions. These two functions can be tried for modelling purposes, if one has a time-series. The methods can be used where a history of call connections is available.

Visick (1980) allows part of the matrix to be projected by a fixed percentage and Kruithof's method to be used on the rest.

David and Pack (1980) report the use of an improved single Kalman filter model, known as the *Sequential Projection Algorithm*, which performs better than existing methods of projection. Saito (1986) also uses a Kalman filter model.

---

[1] Introducing time as another variable just complicates the issue by calling for another variable to be controlled. One basic idea of this thesis is that growth models can be derived mathematically from static models. See equations 14, 17, 19-23 in this chapter and equations 14, 15 in Kennedy (1992).

Time-series forecasting, as documented by Levenbach and Cleary (1984), is one solution that can follow the progress of each node over an extended period. In practice, the constant addition and subtraction of nodes and rearrangement of regions make the historical approach difficult.[1]

2.7 CONCLUSION TO THE LITERATURE REVIEW

The models reviewed in this chapter reveal disagreement among scholars and practitioners as to the number and type of endogenous variables to bring into a call connection model. Endogenous factors considered include the sum of all calls in the network, S; the size of the origin and the destination, O_i and T_j and the growth of each; the distance between the regions, D_{ij} and measurements of N_{ij} gathered previously.

The gravity and growth models have been shown to be flawed when held up to the scrutiny of probability and dimensional analysis. This leaves the former Bell System growth model, equation 17, as a suitable candidate. Chapter five will look closely at the theoretical and empirical basis for this equation.

The next chapter looks for sources of *exogenous* variables and seeks to establish their relationship with calls.

---

[1] The precaution of collecting data a year later was taken, but the data was not processed. Processing would be equivalent to replicating the whole family of experiments. The difficulty is that it will take almost as long as the first run, and will only yield one more point in the time series, which is not enough.

CHAPTER 3
ECONOMIC, REGIONAL & INDUSTRIAL FACTORS AFFECTING CALLS

3.1 MODELLING NODAL USAGE

THIS CHAPTER finds measurable factors that are known to influence the production[1] of calls in the X.25 and X.28 Packet Switched Public Data Network (PSPDN).

Eight major factors are at work in the economy, shaping the total demand for the transfer of information. These factors are economic, regional, industrial, political, governmental, institutional, social and technological. They can be lumped together in national studies. For example, the global study of Terranova and Ulian (1989) shows the relationship between Gross National Product and the total number of telex and data terminations. This chapter differs. First, data calls are the dependent variable. Second, the calls are apportioned over all the regions in one country. Third, by making an explanatory model using just the first three factors, the uncontrolled and difficult-to-measure influence of the final five factors are eliminated.

---

[1] Production-consumption theories are not appropriate to communication networks as calls are vectored quantities, unlike goods and money which are scalars. Telecommunication networks differ from electric power networks where power is transmitted through the grid from any power station to the consumer.

3.1.1 ECONOMIC FACTORS

Economic factors affect the production and the distribution of wealth and the personal and corporate well-being of the country. But how does the economy affect the revenue from a network? And how does it affect the resources required to run a network? This chapter looks at factors that affect the call volume — vital to an understanding of network revenue and required resources.

In the economy, finances flow through the channels of commerce. Funds flow from consumer to producer to pay for goods and services (products). Economists draw a cordon around the country to enable them to contain the hosts of everyday economic activities. They draw this cordon at the final point of sale of products. *Final* means that the end-user (consumer) buys the products in their final form. In this way, economists avoid the sin of double or multiple counting of transactions. Gross Domestic Product (GDP) is *the total value of finished goods and services in a country*. It is a practical measure of the total industry in the domestic economy. Two factors contribute equally to GDP: the industries in the country and the regions of the country.

Not every industry contributes equally to the GDP. Neither does every region. Nor do all industries contribute equally to the network provider's revenue.

3.1.2 REGIONAL FACTORS

Not every region makes an equal contribution to the economy. Some regions are more heavily populated. Others are more productive. Still others are blessed with better resources.

3.1.3 INDUSTRIAL FACTORS

Industries do not contribute equally to the economy either. In a particular country, some industries have better representation than others. Some industries are more efficient at producing financial returns or creating wealth for the country.

3.2 HYPOTHESIS ONE

The present work seeks to form a model that considers the main factors causing calls. The task is to find independent variables that are both logical (significant) and available (measurable). A review of possible sources of network and economic information follows. The closed system studied is the South African packet network and the South African economy.

This chapter explores the first hypothesis that regional data call volumes are proportional to weighted economic activity in the region.

3.3 INDICATORS OF DATA CALL VOLUMES

The call is the fundamental unit common to all telecommunication networks. So, the study employs and explains call volumes, rather than call duration or the number of segments carried.

The problem of providing telecommunication equipment can be divided into two phases. The first phase, considered here, is to establish the number of call attempts. The second phase, not considered here, is to determine the quantity of equipment from the number of calls and the statistical distribution of their lengths.

A single forecast figure for total offered calls is inadequate because a communication network must respond to the need to interconnect any one customer with any other. This chapter excludes the issue of forecasting S, the total number of calls in the network. Instead it concentrates on modelling the geographical apportionment of these calls.

Let N_{ij} be the number of calls originating at node i and terminating on node j. See fig. 2. The total calls originating from node i is called O_i, and O is the column vector containing the sums of call connections over the m terminating nodes in the network:

$$O_i = \sum_{j=1}^{m} N_{ij}. \qquad (34)$$

Similarly, the total calls terminating at node j is called T_j. T is the row vector containing the sums of call connections over the m originating nodes in the network:

$$T_j = \sum_{i=1}^{m} N_{ij}. \qquad (35)$$

In addition, it is expedient to define the average usage of node r as

$$U_r = (O_r + T_r)/2. \qquad (36)$$

This will allow the regions to be ranked uniquely.

Like the total number of calls in the network, nodal usage is subject to seasonal variations and business cycles. In addition, nodal usage is subject to variations from the eight major economic factors listed at the start of this chapter. With a suitable choice of measuring period, these variations in the number of calls can be eliminated, leaving the base data.

3.4 OBTAINING DATA CALL RECORDS

The X.28 and X.25 services may be partially substituted for each other. Consequently, these calls were combined for the study. The better correlations that resulted vindicated this approach.

Call records were derived from files used for billing purposes. As metering information is subject to financial audits and is used to charge the customer, the accuracy of this source was expected to be impeccable. (However, cases of over-metering were identified and deleted from the data.) Measurements spanning a full month were used to ensure that the full spectrum of traffic (including that at the beginning or end of the month or week) was captured. Records for one month were regarded as forming a suitable sample, since the sampling proportion was very high (100%). It is unlikely that wide swings in the relative traffic levels of each region will occur from month to month. Since calls were to be the dependent variable, it was necessary to select one of the following two options:

- call volume per year, quarter, month, week or day;

- call intensity per busy hour, minute or second.

The latter option is independent of the calling profile and tariff changes during the day, but demands that the busy period be first identified. As time and date of call were not available in the extract of the billing file supplied, the former option was used.

3.5 REGIONAL ACTIVITY INDICATORS

A review of possible sources of economic data now follows. The number of calls generated can be explained using an item from the following list as input variable:

- The Central Statistical Service (CSS) compiles and publishes official figures (CSS 1989b). This includes demographic figures. See for example CSS (1985). Population was soon found to be a poor predictor.
- The CSS also publishes official figures of the regional breakdown of GDP. See for example CSS (1989b).
- Revenue derived by the telecommunication administration.

The list is in increasing order of accuracy (which coincides with increasing cost of collection).

3.5.1 GROSS REGIONAL PRODUCT

Gross Regional Product (GRP) has not been formally defined in the literature and is defined here as *the final remuneration received by land, labour, capital and entrepreneurship for its participation within a region*. GRP is preferable to demographic data as not all members of the population are equally active in the economy. GRP can be assumed to be a better predictor of calls than other statistics such as population, salaries, or number of customers. This assumption is based on the self-evident observation that not all members of the population (even salaried members) are customers of the data networks. Even less are active customers of the networks. So, GRP was extracted from official figures obtained from the CSS and used as the main explanatory variable. Although there exists a gap of six years between the date of the official figures and the study data, this does not invalidate the study.

The disadvantages of GDP (and by extension GRP) are well-known. Mohr (1988) points out: Significant activities go unreported. Non-market production has been estimated. The data is subject to revisions and is not a full measure of well-being. Nevertheless, GRP reveals itself to be a powerful means of explaining calls.

Also, there is not a perfect mapping of the regions of the CSS, the X.25 nearest node areas and the X.28 nearest multiplexer areas. However, the data network switches may be aggregated into approximations of the CSS or other regions. Many advantages accrue: Zoning changes caused by additional nodes or access points may easily be accommodated. Differential growth of each region is incorporated in a natural fashion. Standardized regions are used, easing communication with other organizations.

3.6 PROCEDURE

In this country, a magisterial district is the smallest area of legal jurisdiction. The GDP components for each magisterial district and industry were obtained from the CSS (1989b) and entered into a computer. See the appendix. Totals were checked against those printed in the source. Standard codes for each magisterial district (including check digits) were entered from the CSS (1989c) to allow subtotals to be checked. Allowance was made for the rounding off to the nearest kilorand followed by the CSS.

Data for the entire EDXP network was extracted in a process parallel to the billing run. Certain fields that were not required were omitted. For example, count per tariff period and per message class were excluded. A preliminary analysis revealed that the tape contained records of extraneous events that masked the true traffic patterns. For example, one record revealed that 1,8 million calls attempts had been made from one address to another during the month. This was traced to faulty programming of a packet assembler-disassembler. Table 8 gives the logical criteria that were used to ensure that only valid data were down-loaded from the billing tape. The filtered, grouped data appear in table 1. Ignore, for the moment, the last column in the table.

A standard test was used to find R^2, the square of the correlation coefficient, known as the *co-efficient of determination*. This gives the degree of correlation between model and measurements. See fig. 3.

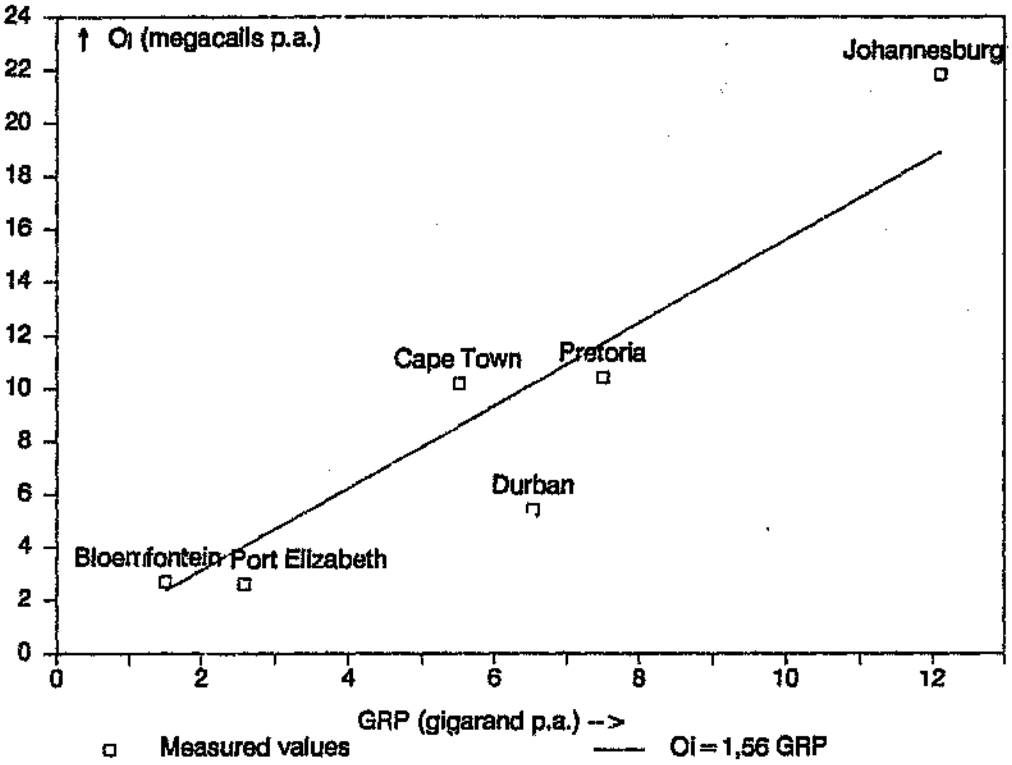

Figure 3. Originating call volume versus unweighted GRIP

Table 1. Unweighted and weighted GRIP

Region	GRP (gigarand p.a.)	O_i (megacalls p.a.)	GRIP (gigarand p.a.)
Johannesburg	12,114	21,816	11,798
Pretoria	7,493	10,428	5,841
Cape Town	5,506	10,188	5,608
Durban	6,530	5,424	4,896
Bloemfontein	1,496	2,688	1,383
Port Elizabeth	2,580	2,592	1,917

From theoretical considerations, and from an inspection of the data, the usage must be zero when GRP is zero. Therefore a zero intercept was assumed during the fitting of a straight line to the data, and in calculating R^2. This precaution is recommended by Kvalseth (1985).

STATISTICAL ASSUMPTIONS

The standard index R^2 indicates how well experimental data fits an equation. (R^2 only applies if the relationship is linear. If it is not, it must first be linearized, e. g. by taking logarithms.) A figure of 1,0 indicates perfect correlation and 0,0 indicates no correlation. Values between these extremes indicate the degree of correlation. According to Orvis (1987), $R^2 > 0,9$ corresponds to a "good fit". However, Catanese (1972) maintains that $R^2 > 0,87$ is "quite high for many urban problems". To confuse the issue further, Wheelwright and Makridakis (1985) indicate that $R^2 > 0,78$ corresponds to "good results". For this thesis, $R^2 > 0,80$ is assumed "good" and $R^2 > 0,75$ is assumed "satisfactory".

3.7 PRELIMINARY RESULT

The analysis of correlation between usage and GRP gave $R^2 = 0,839$ which is good, being greater than the 0,80 figure chosen as a threshold.

3.8 ANALYSIS AND DISCUSSION

Showing that two variables are correlated is a necessary step but is not sufficient to show that one variable causes the other. However, if there is a logical link between the two variables, it is reasonable to say that one variable causes the other.

It appears logical that in areas of greater economic activity there will be a greater demand for data networks or vice versa. This is confirmed by the correlation results.

The division of the country into the regions used was a purely arbitrary grouping. The choice made was only for convenience of data reduction. Thus it is reasonable to generalize that if the country is divided into any other arbitrary groups, data calls and economic activity will still be found to be closely related.

3.9 IMPROVING THE MODEL

Can the accuracy of the model be enhanced?

Clearly, calls are not simply caused by the gross economic activity. Calls are more closely tied to the mix of active industries and the relative intensity with which these industries use the public data networks.

Some factors that could indicate the mix of industries were needed. A strategy was devised to use weighted regional and industrial economic statistics as a predictor of calls. By weighting the components of GDP, it is possible to model the number of calls from each region more closely.

First, however, it is necessary to examine the constituents of GDP and examine revenue from the data network as a source of information.

3.9.1 GROSS REGIONAL AND INDUSTRIAL PRODUCT

Industries are classified by the CSS (1988) at the highest level into one of ten major categories: agriculture, mining, manufacturing, utilities, construction, trade, communication, financing, services and government. Each industry is composed of users whose needs are roughly homogeneous. That is, there is competition within the industries, but not between the industries. The size and needs of each industry give the relative profitability of providing service to that industry. Each *major Standard Industry Code* (c) and its abbreviation as used in the appendix appear in Table 2. The subscript c is used as a place holder for the columns containing different industries. The letter chosen serves as a reminder that it is derived from the industry or commerce code. The table also shows a final logical grouping of the major industries.

It is possible to regard GDP as an aggregation not only over the m regions, r of the country, but also over the ten major Standard Industry Codes. That is,

$$GDP = \sum_{r=1}^{m} \sum_{c=1}^{10} GRIP_{rc}. \qquad (37)$$

$GRIP_{rc}$ has not been named in the literature, but will now be termed the *Gross Regional and Industrial Product*. GRIP is defined here as *the final remuneration received by land, labour, capital and entrepreneurship by an industry for its participation within a region*.

Table 2. Abbreviations of major standard industries

c	Major industry	Abbreviation	Further grouping
1	Agriculture	Farm	*Primary Industries*
2	Mining	Mine	
3	Manufacturing	Make	*Secondary Industries*
4	Electricity	Supply	
5	Construction	Build	
6	Trade	Trade	*Service Industries*
7	Transport	Move	
8	Financing	Bank	
9	Community Services	Serve	
10	Government	Gover	*Non-profit Seeking Industries*
0	Others	Produc	

The GDP components for each magisterial district and industry were obtained from the CSS (1989b) and entered into a computer. The data appear in the appendix. Horizontal and vertical totals were checked against those printed in the source, making allowance for the rounding off to the nearest KiloRand followed by the CSS. Standard codes for each magisterial district (including check digits) were entered from the CSS (1989c) to allow subtotals to be checked. The data were loaded into the program NetCalc 1.06 written by Kennedy (1991a) so that simultaneous grouping of rows and columns could take place.

Each industry of the economy contributes separately to the revenue from data services. In practice, it was found that the agricultural, mining, manufacturing and construction industries were not represented, and were excluded. This leads to a five industry model, requiring less data entry and storage.

Industry Revenue (NR_c) has not been named in the literature and is defined here as *the annual income received for the provision of a network service to a major industry, c.*

Network Revenue (NR) has not been named in the literature and is defined here as *the annual revenue earned from all industries by a country-wide network.* It is aggregated as:

$$NR = \sum_{c=1}^{m} NR_c. \tag{38}$$

The concept of revenue from a particular industry will be used next to quantify the relative importance of each industry to the data network provider.

3.9.2 DATA NETWORK INTENSIVE RATIO

The predominant users of the South African public data networks are business clients. The penetration of the residential market is very limited, such customers using their service for business purposes. Business users may be classified according to industries.

Some industries are introspective by nature having little need for country-wide data communication. Such industries include the agricultural industry. Other industries must use country-wide data communications to survive: the very existence of their business relies on it.

Each industry uses country-wide data communications to varying degrees. Thus each industry does not play an equal role in contributing to the network revenue. It becomes necessary to introduce a *weighting factor*, here named the *Data Network Intensive ratio* $DNI_c = DNI_1 \dots DNI_{10}$ for each industry of the economy.

Here the weight, (DNI_c), for a particular industry is the industry revenue, NR_c, from the major industry, c, expressed as a fraction of the Gross Industrial Product of that industry:

$$DNI_c = NR_c / GIP_c. \tag{39}$$

Since each major industry's use of the network is at a different level, it is necessary to apply the industry weighting to each GRIP to account for the profile of the customers.

3.10 ANALYSIS OF NETWORK REVENUE

Institutions using the data networks were classed into one of two account types according to whether their account for public data services in May 1987 was above or below R100 000. See table 3. The accounts include fixed and usage sensitive charges, but exclude telephone, telex and videotex charges.

Table 3. Analysis of data accounts

ACCOUNT	NUMBER	AMOUNT (R)
Major (>R100 000 p.m.)	18	5 244 636
Minor (<100 000 p.m.)	3 146	4 500 267
NR	3 164	9 744 903

Although the top 18 major account holders number only 0,6% of the institutions, the income from them formed about 54% of the service provider's income. Correspondingly, one can expect that most calls come from these few users. A scrutiny of them is merited.

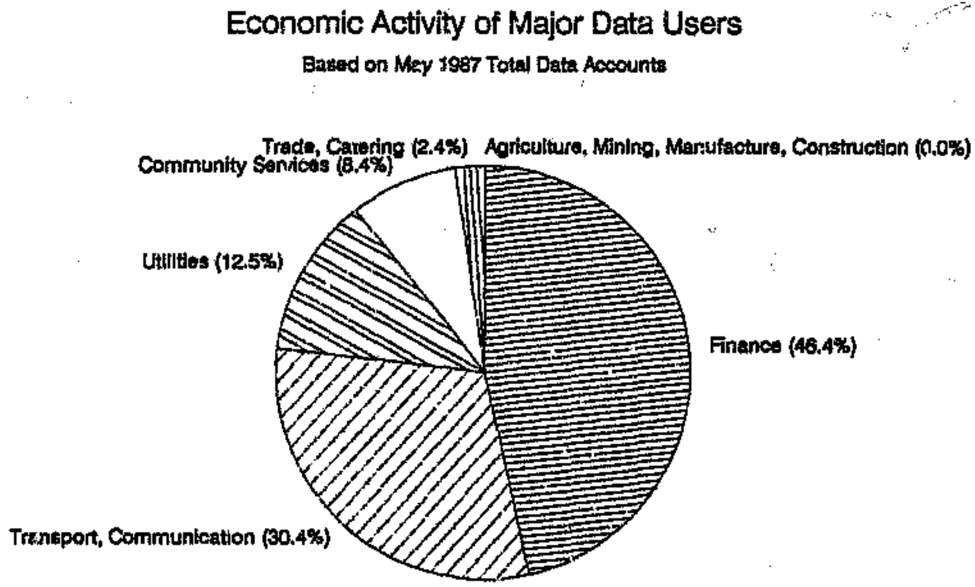

Figure 4. Data network customer profile

3.10.1 THE MARKET SEGMENTATION

The specific mix of data network users was established. Fig. 4 shows the particular mix of data users being served in this country. It was derived from the May 1987 revenue census. The agricultural, mining, manufacturing and construction industries are effectively excluded.

3.10.2 DATA NETWORK INTENSIVE RATIO

The major users of data communication services were grouped according to the CSS (1988) Standard Industrial Classification (SIC) of major industries. A new concept, the *Data Network Intensive (DNI)* ratio was introduced. It is calculated as follows:

$$DNI_c = \frac{\text{12 × monthly data-account of major data-netwotk users}}{\text{gross annual insustrial product of major industry}} \qquad (40)$$

A convenient unit for expressing DNI_c is a unit of currency (Rands) per thousand units of currency, each expressed per year. The numerator was measured in Rands (for May 1987) and the denominator was measured in KiloRands (for 1984). DNI shows the degree to which that major industry is involved in public data networking. The higher the figure, the more is paid to the administration for data services per unit production. As can be expected, Table 4 shows that industries that are perforce actively involved in the transfer of information around the country have higher DNI ratios.

Table 4. Data Network Intensive ratio and industry (major accounts only)

c	DNI_c	INDUSTRY	
8	2,72	Financial, insurance, real estate & business services	
7	1,78	Transport, storage and communication	
4	0,73	Electricity, gas and water	
9	0,49	Community, social and personal services	
6	0,14	Wholesale & retail trade, catering & accommodation services	
0	0,00	Activities not adequately defined	)
1	0,00	Agriculture, hunting, forestry and fishing	) NO
2	0,00	Mining and quarrying	) MAJOR
3	0,00	Manufacturing	) ACCOUNTS
5	0,00	Construction	)

3.10.3 FINAL RESULTS

The annual usage of the packet networks as a function of GRIP is plotted in fig. 5, with the best straight line fitted to go through the origin. In the figure, the horizontal axis shows the weighted GRIP. The vertical axis shows the originating calls per year. If originating calls are proportional to weighted GRIP, then the points would fall on the straight line graph. The closeness of the plotted points to the straight line graph shows how accurate the hypothesis is. The correlation analysis gave $R^2 = 0{,}954$ which is excellent, being better than the 0,856 figure achieved without weighting GRIP. The best fit between calls and weighted GRIP was:

$$O_i = 1{,}75 \times W_r \text{, where} \qquad (41)$$

$$W_r = \text{Weighted GRIP} = \sum_{c=1}^{10} \text{GRIP DNI}_c \text{, and} \qquad (42)$$

c = major Standard Industrial Classification, O_i = total measured calls originating p.a.,
GRIP = Gross Regional and Industrial Product in KiloRands p.a., and
DNI_c = Data Network Intensive ratio for the industry in dimensionless units.

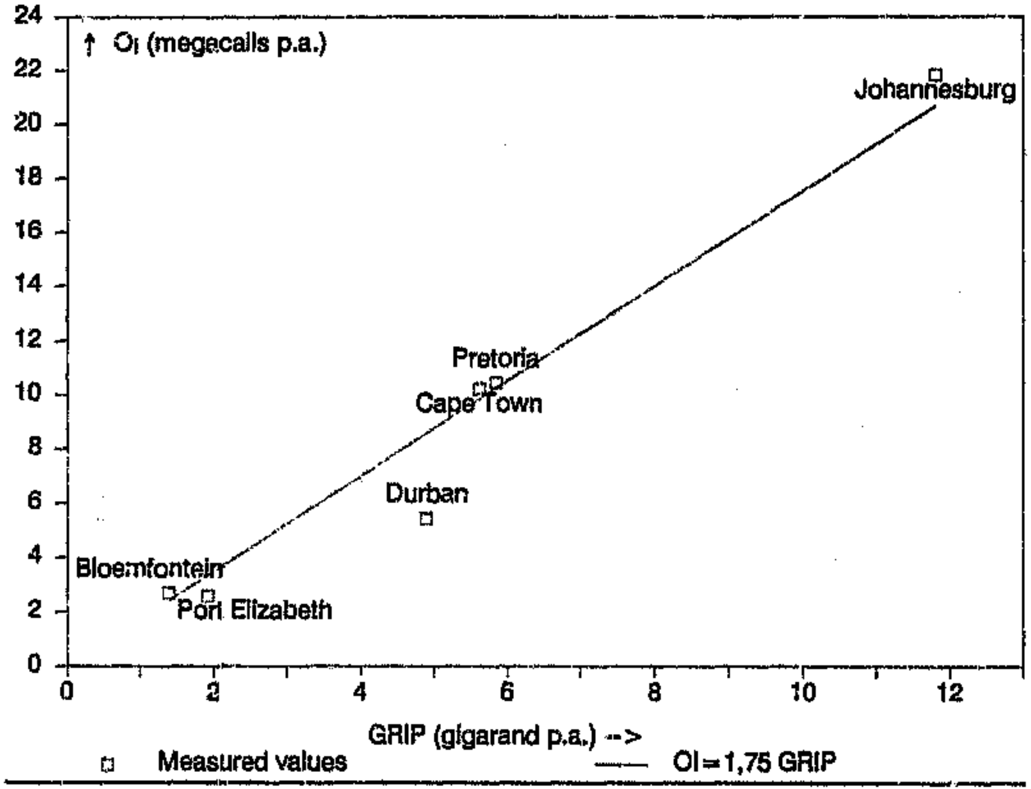

Figure 5. Originating call volume versus weighted GRIP

3.11 INTERPRETATION OF RESULTS

It is likely that an even better R^2 could be displayed if weighting factors were to be determined from the PSPDN call records themselves. That is, if each user was classified according to industry and the weighting factors determined from the number of calls measured from each industry. These steps were not performed as such a method would entail too much work for the benefits derived. The method adopted provides an adequate model for planning purposes.

The meaning underlying the relationship found above is that calls are caused by the economic activity or vice versa.

Clearly calls are not simply "caused" by the general industrial activity in that region, but are more closely tied to the mix of active industries and the relative intensity with which the mix uses the networks. In this country, every KiloRand of weighted production in an area caused 1,75 calls to originate in the PSPDN (and the same number to terminate). Expressed the other way around: 1,75 calls originating were needed to support the weighted production of goods or services worth one KiloRand.

3.12 CONCLUSION

Conclusion 1: The number of calls made by PSPDN customers is proportional to the Gross Regional and Industrial Product of the region.

Conclusion 2: An improved model is obtained by weighting the most recent Gross Regional and Industrial Product for the (10) industries according to the Data Network Intensive ratios from the major network users in the (5) most important of the major industries.

Conclusion 3: The financial industry is the major customer. This category includes all financial institutions, insurance or real estate companies, business services and machinery or equipment renting and leasing companies.

3.13 OTHER RESULTS

Because of the limited capacity of each packet switch, the load on each node is divided among several packet switches. Interestingly, the sixteen switches serving the top six nodes carried 4 399 kilocalls per month but "the rest", the remaining thirteen switches, carried only 75 kilocalls per month.

3.14 SUMMARY

Hypothesis one was fully supported by the results. It was found that 95,4% of all variation in the usage of a node is accounted for by a hypothesis of a linear relation of usage with weighted Gross Regional and Industrial Product. Data calls are thus satisfactorily correlated with regional and industrial economic activity.

Further, it was found that weighted GRIP provided a better correlation with the observed packet switched call volumes than did *unweighted GRIP* (GRP). Calls and economic activity are thus closely related. Kennedy (1991c) reports similar results if X.28 calls are excluded. Economic activity could conceivably continue (somewhat hampered) without calls through the data networks. However, the converse is not true: Calls could not continue without economic activity. The inescapable conclusion is that economic activity causes calls.

The next chapter looks at a new hypothesis relating to the ranking of economic activity.

CHAPTER 4
APPLICATION OF THE RANK-SIZE RULE

4.1 INTRODUCING THE RANK-SIZE RULE

IN PLANNING a network for a country, a planner may be unsure of the absolute sizes of the centres, but be able to estimate their relative ranks. In such cases, the rank-size rule is a useful planning tool, which is little known outside the field of regional economics. The formula furnishes an acceptable approximation to regional data though it appears to lack a theoretical foundation. The rule states that *the population of a city is inversely proportional to its rank*. That is, the second largest city has a population half that of the largest city; the third largest city has a population one third that of the largest city, and so on. The rule requires that a self-contained, large study area be selected, note Hoover *et al.* (1985). The accuracy of the rule relies upon a reasonable grouping of cities with associated centres. The rule is a surprisingly simple description of perplexingly complex phenomena. The rank-size rule captures the mechanism by which centres compete: the larger the centre, the greater the growth. Up to now it has not been shown whether the rank-size rule can be applied to units of size other than population.

This chapter extends the rank-size rule to non-demographic data. The body of the chapter has three main sections, corresponding to the variables studied: Demographic data, Gross National Product, and Gross Regional Product with Gross Regional and Industrial Product.

4.2 HYPOTHESIS TWO

In this chapter, attention is temporarily diverted from studying data calls to the topic of demographics and to regional and industrial economics. The chapter asks: What are the patterns behind population or Gross Regional and Industrial Product?

This chapter tests the second hypothesis that population and weighted and unweighted GRIP in this country follow the rank-size rule well enough for modelling purposes.

4.3 DEMOGRAPHIC DATA AND THE RANK-SIZE RULE

This section tests how closely demographic data comply with the rank-size rule.

The official population of each centre in the country was obtained from the Central Statistical Service CSS (1985). Although the census suffered from a variable undercount, it is the only official population count available.

The figure for Johannesburg was formed by totalling the population of the city itself and the towns of the Witwatersrand (Roodepoort and Germiston). The cities were then sorted and ranked. The population and ranks appear in table 5.

Table 5. Population and rank of cities in South Africa

City	Rank	Population	Model
		(millions 1985)	
Johannesburg	1	2,953	2,953
Cape Town	2	1,544	1,477
Durban	3	0,982	0,984
Pretoria	4	0,822	0,738
Port Elizabeth	5	0,651	0,591
Vereeniging	6	0,461	0,492
Bloemfontein	7	0,233	0,422
Pietersburg	8	0,205	0,369
Pietermaritzburg	9	0,192	0,328
East London	10	0,185	0,295
Kimberley	11	0,149	0,268

4.3.1 POPULATION POSTULATE

In this country, P, the population of the r^{th} city, is postulated to follow the model

$$P_r = P_1/r, \tag{43}$$

where r is the *rank* of the city and P_1 is the number of people in the most populous area.

4.3.2 PROCESSING OF POPULATION RANKS

The agreement of the population model with measurements was gauged using R^2.

4.3.3 POPULATION RESULTS

R^2 was 0,981. The closeness of this result to unity shows the excellent agreement of the model with the measurements. A graph of the modelled values against the published values appears in fig. 6. The horizontal axis gives the population as derived from the CSS. The vertical

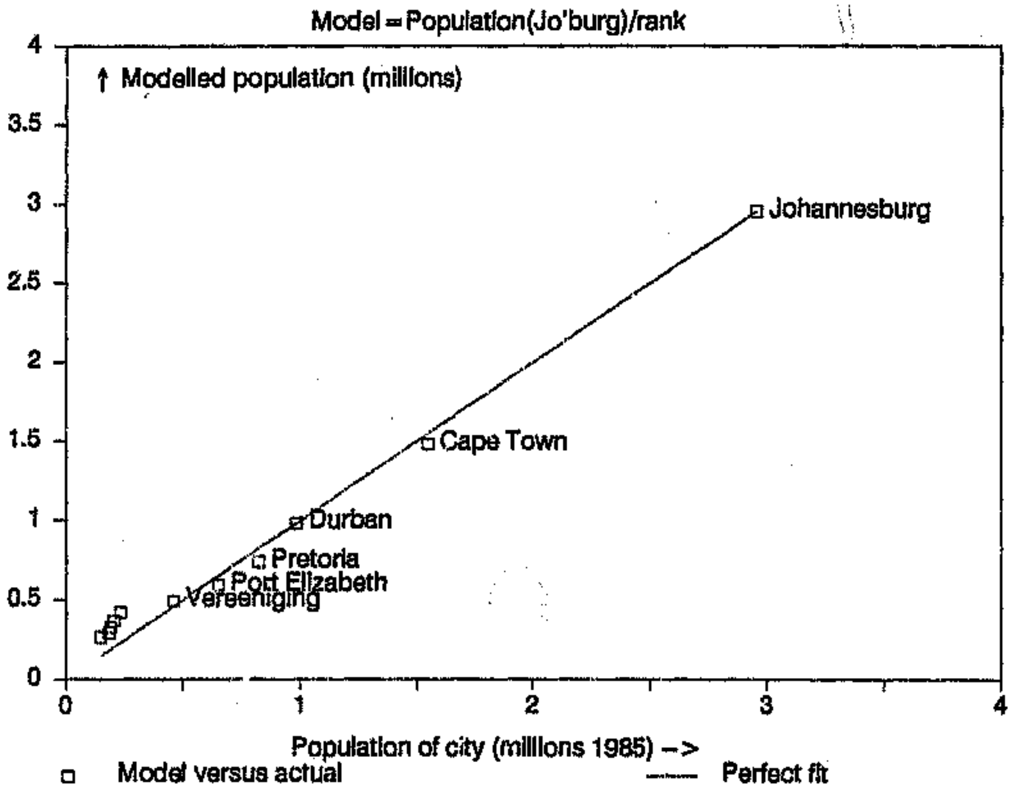

Figure 6. Rank-size model for South African population

axis gives the values modelled by the rank-size rule. The straight line shows a perfect fit of model with reality. The square symbols are plots of the modelled values against the published values. These squares lie close to the straight line graph that corresponds to a perfect fit of model with reality.

4.3.4 POPULATION CONCLUSION

The population of the cities of South Africa follows the rank-size rule, with 98,1% of the variations being accounted for by [illegible] model. This principle is not new to regional economists. However, the quantification of the strength of the relationship and what follows below is new.

4.4 GROSS NATIONAL PRODUCT

The results achieved above for demographic data lead to the question: does the rank-size rule also apply to economic activity? [illegible]ere is a correlation between GNP and population and GNP or GRIP, one might expe[illegible] and GRIP to follow the rank-size rule.

To answer the question on national scales, it is appropriate to emphasize the national character of production. *Gross National Product* (GNP) is the Gross Domestic Product less the net return on foreign investment. This section examines GNP to see if it also conforms with the rank-size rule.

4.4.1 SOURCE OF GNP DATA

Kurian (1979) sorts, ranks and lists global statististics. In particular, he tabulates sorted and ranked statistics of countries of the world. He ranks 119 countries according to their GNP. The GNP and ranks appear in table 5. For reasons of economy only, the series was truncated after South Africa was reached. The bottom end of the scale hardly affects R^2.

4.4.2 GNP POSTULATE

The GNP of the r^{th} country was postulated to follow the model

$$GNP_r = GNP_1/r, \tag{44}$$

where r is the rank of the country and GNP_1 is the GNP of the country producing the most.

4.4.3 PROCESSING OF GNP RANKS

R^2 was calculated to find the agreement of the GNP rank-size model with measurements.

Table 6. GNP and rank of countries

Country	Rank	GNP (teradollars '76)‡	Model
United States	1	1,698	1,698
Soviet Union	2	0,708	0,849
Japan	3	0,553	0,566
West Germany	4	0,458	0,425
France	5	0,347	0,340
China	6	0,343	0,283
United Kingdom	7	0,225	0,243
Canada	8	0,174	0,212
Italy	9	0,171	0,189
Brazil	10	0,126	0,170
Spain	11	0,104	0,154
Poland	12	0,098	0,142
India	13	0,096	0,131
Netherlands	14	0,085	0,121
Australia	15	0,083	0,113
Sweden	16	0,071	0,106
East Germany	17	0,071	0,100
Mexico	18	0,068	0,094
Belgium	19	0,067	0,089
Iran	20	0,066	0,085
Czechoslovakia	21	0,057	0,081
Switzerland	22	0,057	0,077
Turkey	23	0,041	0,074
Austria	24	0,040	0,071
Argentina	25	0,040	0,068
Saudi Arabia	26	0,039	0,065
Denmark	27	0,038	0,063
Yugoslavia	28	0,036	0,061
South Africa	29	0,035	0,059

4.4.4 GNP RESULTS

The value of R^2 was 0,986. This again shows an excellent agreement between model and measurements. The agreement can be seen in fig. 7. The horizontal axis gives the GNP as listed by Kurian (1979). The vertical axis gives the values modelled by the rank-size rule. The

---

‡ kilo = 10^3; mega = 10^6; giga = 10^9; tera = 10^{12}.

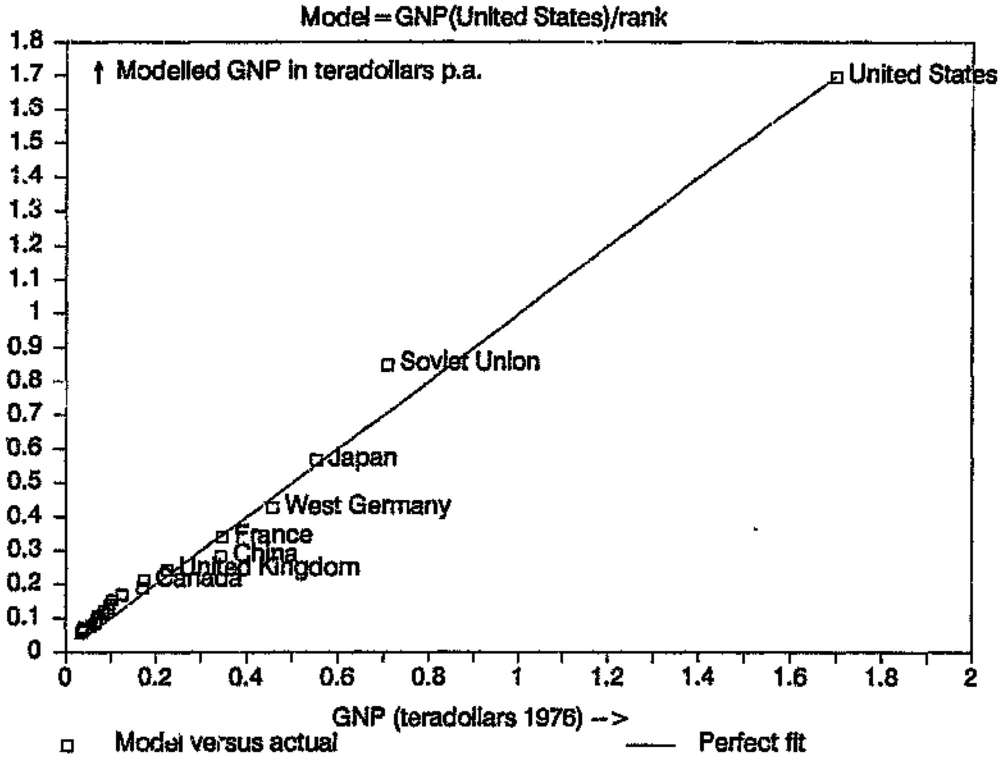

Figure 7. Rank-size model for Gross National Product

squares are plots of the modelled values against the actual values. Again, the squares lie close to the straight line graph that corresponds to a perfect fit of model with reality.

4.4.5 GNP CONCLUSION

The GNP of countries of the world follows the rank-size rule, with 98,6% of the variations being accounted for by the model.

4.5 GROSS REGIONAL & INDUSTRIAL PRODUCT RANK

Knowing now that Gross National Product conforms with the rank-size rule, this section examines the Gross Regional and Industrial Product of regions of this country to see if unweighted and weighted GRIP conform with the rank-size rule.

4.5.1 SOURCE OF GRIP DATA

Official CSS (1989b) data for the components of GDP were obtained and weighted as described in the previous chapter. The unweighted and weighted GRIP and ranks appear in Tables 7 and 8. The different rankings (e.g., for Cape Town) are because the city contains a large proportion of the target market (e.g., financial institutions).

4.5.2 GRIP POSTULATE

The weighted or unweighted GRIP of the r^{th} region in this country was postulated to follow the model

$$GRIP = GRIP_1/r, \qquad (45)$$

where r is the *rank of the region* and $GRIP_1$ is respectively the weighted or unweighted GRIP of the region with the largest GRIP.

4.5.3 PROCESSING OF GRIP RANKS

R^2 was calculated to find the agreement of the weighted or unweighted GRIP rank-size model with measurements.

4.5.4 GRIP RESULTS

The horizontal axis of fig. 8 gives the unweighted GRIP and the horizontal axis of fig. 9 gives the weighted GRIP as found in the previous chapter. The vertical axes give the values modelled by the rank-size rule. The squares are plots of the modelled values against the actual values. These squares are seen to lie close to the straight line graphs that correspond to a perfect fit of model with reality. The values of R^2 were 0,987 and 0,954 for unweighted and weighted GRIP respectively. Again, this shows an excellent agreement between model and measurements.

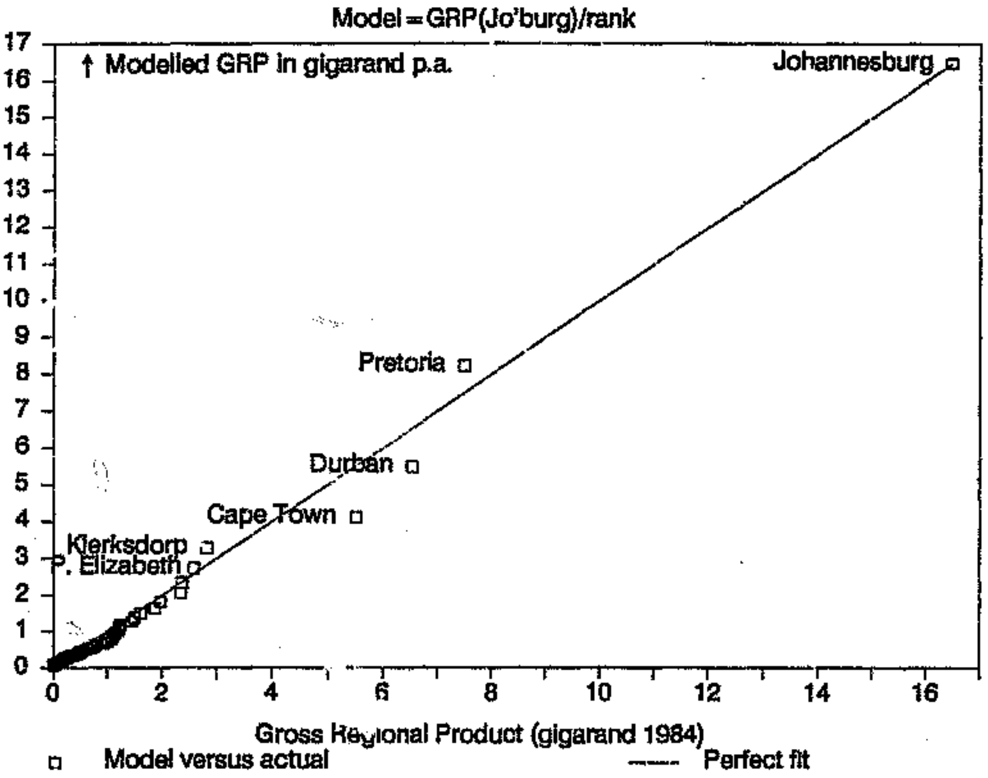

Figure 8. Rank-size model for unweighted GRIP

Table 7. Unweighted GRIP and rank of regions

Region	Rank	GRP (gigarand p.a.)	Model (gigarand p.a.)
Johannesburg	1	16,467	16,467
Pretoria	2	7,493	8,234
Durban	3	6,530	5,489
Cape Town	4	5,506	4,117
Klerksdorp	5	2,815	3,293
Port Elizabeth	6	2,580	2,745
Oberholzer	7	2,370	2,352
Highveld Ridge	8	2,332	2,058
Welkom	9	1,969	1,830

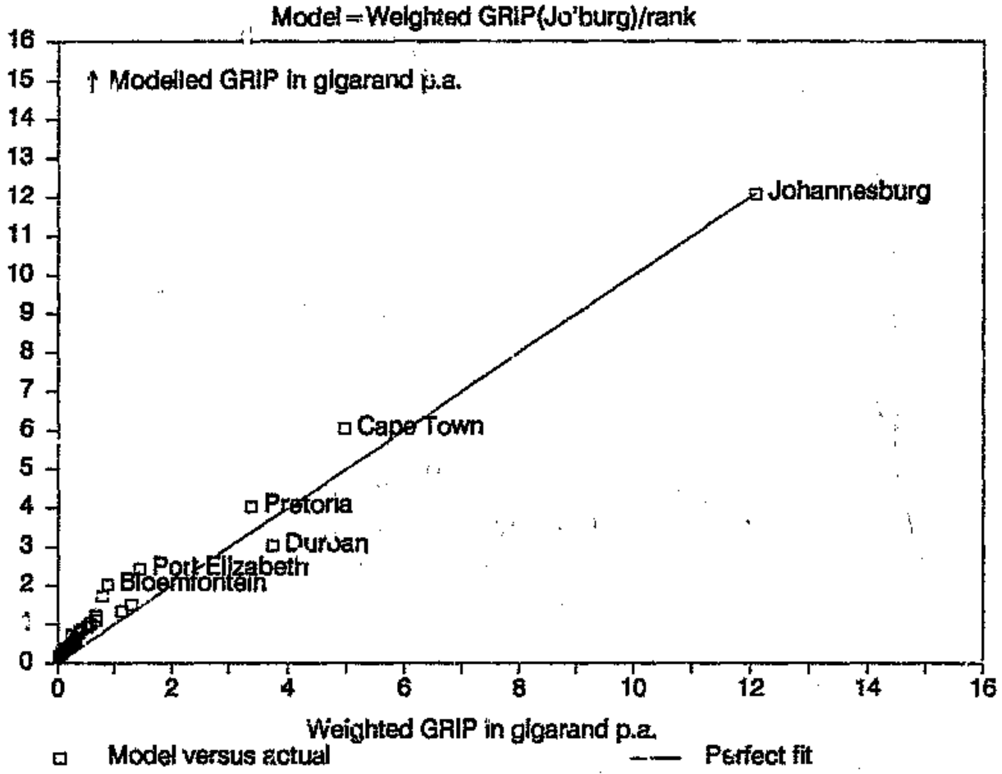

Figure 9. Rank-size model for weighted GRIP

Table 8. Weighted GRIP and rank of regions

Region	Rank	GRIP (gigarand p.a.)	Model
Johannesburg	1	12,077	12,077
Cape Town	2	4,960	6,038
Pretoria	3	3,374	4,026
Durban	4	3,739	3,019
Port Elizabeth	5	1,438	2,415
Bloemfontein	6	0,871	2,013
Oberholzer	7	0,784	1,725
Witbank	8	1,301	1,510
Bethal	9	1,134	1,342
East London	10	0,678	1,208
Vereeniging	11	0,666	1,098
Pietermaritzburg	12	0,580	1,006
Paarl	13	0,509	0,929
Phalaborwa	14	0,418	0,863
Kroonstad	15	0,371	0,805
Kimberley	16	0,366	0,755

4.5.5 GRIP CONCLUSION

The weighted and the unweighted GRIP of the regions of the country follow the rank-size rule, with respectively 95,4% and 98,7% of the variations being accounted for by the model.

4.6 CONCLUSION

The population of cities in this country follows the rank-size rule, with 98,1% of the variations being explained. The rank-size rule explains 98,6% of the variations that occur in GNP. It can also account for 95,4% of the variations that occur in unweighted GRIP and 98,7% of the variations that occur in GRP in this country. Thus if the sizes of centres are unknown, their ranking may be used instead as an explanatory variable, with only a little loss of information.

4.7 DISCUSSION OF THE RANK-SIZE RULE

The rank-size formula furnishes an excellent estimate of nodal size, though it lacks a theoretical foundation. The reasons for this success are fourfold: The function is forced to go through the first data point. Second, its success lies in its gross simplifying approach. Third, it yields the non-linearity that matches reality. Fourth, since both are sorted, there is an axiomatic auto-correlation between rank and size that encourages the rule to work.

4.8 SUMMARY

This chapter has shown that GNP obeys the rank-size rule. Subsequently, this chapter has shown for South Africa that the rank-size rule applies not only to population but also to weighted and unweighted GRIP. So, where population or economic figures are lacking, it is possible to model them, providing one can estimate the relative ranks of the regions.

The importance of this result to the thesis is that an estimate of the size of a node can be obtained algorithmically from the rank of the node. Later this thesis will extend the result found for nodes to links. Meanwhile, the next chapter examines the size of links from a probability viewpoint.

CHAPTER 5
ANALYSIS OF FORMER BELL SYSTEM GROWTH MODEL

5.1 INTRODUCTION TO INVESTIGATION

THIS CHAPTER provides a first estimate of the number of calls within and between any regions in a packet network. The most suitable model from chapter two is used to forecast calls between regions. The basis of the former Bell System growth equation was selected as the most promising model. The shortcoming of the model is shown theoretically and empirically to be a result of the impossibility of meeting the implicit assumption of statistical independence.

The chapter is structured classically into the hypothesis, method, results, discussion and conclusion.

5.2 HYPOTHESIS THREE

A null hypothesis will be used to disprove the theory that variations between the observed and predicted call connections can be explained only in terms of pure chance. The probability P(ij) of a call arising between two regions, i and j, is (null) hypothesized to be equal to the joint probability of a call originating in region i and terminating in region j.

This chapter considers the third hypothesis that the origin of calls is independent of their destination. In other words, statistical independence of origin and destination exists in the network. The null hypothesis is:

$$P(ij) = P(i)\,P(j). \tag{46}$$

This *null hypothesis* assumes that call origins and destinations are chosen independently. Empirical data will be used subsequently to show that the assumption is invalid.

From the definition of probability, equation 46 means in terms of relative frequencies that

$$N_{ij} / S = (O_i / S)\,(T_j / S) \tag{47}$$

$$\text{i.e., } N_{ij} = O_i\,T_j / S. \tag{48}$$

Equation 48 will now be developed into a growth equation (i.e., one in which time features). At a future point in time (denoted by an asterisk),

$$N_{ij}^* = O_i^* T_j^* / S^*. \tag{49}$$

Dividing equation 49 by equation 48, and using the definition of growth ratio, $G = N^*/N$ gives equation 17 in chapter two. That is, the theoretical basis for equation 17 is the null hypothesis above.

5.3 METHOD

Raw accounting data were found to be available, but the data had never before been processed to extract valid call connections, N_{ij}. The call records were filtered according to the logical criteria in table 9. This ensured that extraneous factors were excluded and only valid data was downloaded from the billing tape.

Table 9. Logical filtering of call records

Exclude	Download Logic
Lost calls	B_ADDRESS > 0
Faulty Packet Assembler Disassemblers	CALLS < 100000
Switchover faults	DURATION < 21548236
International calls	INTERNAT < 1
Namibian calls	NAMIB < 1
Other category calls	OTHER < 1
Test calls	DURATION > 0
No data interchanged	SEGMENTS > 0

Because of the limited capacity of packet switches, multiple switching units serve certain regions. Thus the data was grouped accordingly. See table 10.

Table 10. Grouping performed

Region	PSPDN Node Numbers
Pretoria	2, 3
Johannesburg	4 - 6, 11, 28, 29, 34
Cape Town	7, 8, 14
Bloemfontein	9
Durban	10, 35
Port Elizabeth	13
Klerksdorp	15
Witbank	16
Newcastle	17
Vereeniging	18
East London	19, 20
Kimberley	23
Welkom	24
George	25
Pietermaritzburg	26, 27
Pietersburg	31
Nelspruit	32
Unused	1, 12, 21, 22, 30, 33

Again, a standard test was used to find R^2, the square of the coefficient of correlation between model and measurements.

5.4 RESULTS

Table 11 lays out the empirical values of N_{ij}.

Table 11. Empirical N_{ij} (kilocalls Sept. 1989)

	FROM \ TO	Jhn	Prt	CpT	Dur	PE	Blm	Rst	$O_i =$
Jhn	Johannesburg	589	506	221	445	18	35	4	1 818
Prt	Pretoria	172	381	88	46	103	66	13	869
CpT	Cape Town	291	146	286	91	23	9	3	849
Dur	Durban	136	134	80	60	8	32	2	452
PE	Port Elizabeth	56	78	57	5	11	3	6	216
Blm	Bloemfontein	36	68	20	9	2	88	1	224
Rst	The rest	9	8	10	6	5	2	6	46
	$T_j =$	1 289	1 321	762	662	170	235	35	S = 4 474

A further test was performed on the data given in table 11, using equation 48 as a model. The chi-square statistic was calculated as:

$$\chi^2 = \sum_i \sum_j (N_{ij} - O_i T_j / S)^2 / (O_i T_j / S) \qquad (50)$$

Equation 50 yielded $\chi^2 = 1\,369$. Using 36 degrees of freedom, this value is larger than the critical value published in tables. (According to Ostle *et al.* (1975), $\chi^2_{0{,}999(36)} = 68{,}0$.) The level of confidence was 99,9%. Using units of kilocalls instead of calls makes the result more certain, and rejection of the null hypothesis more definite.

Thus the null hypothesis is rejected. That is, the assumption behind the null hypothesis is not upheld.

5.5 DISCUSSION

However, the χ^2 test merely proves that there are factors at work other than random factors causing the residual variations in the call connections. There are four of these factors which are:

(1) Because they refer to the same regions, O_i and T_j are strongly correlated.

(2) Also, N_{ij} and N_{ji} are strongly correlated (and should be nearly equal) since these are *complementary call connections.* Yet the measured matrix is asymmetric about the major diagonal, revealing examples where N_{ij} is much greater than N_{ji}, e.g. Johannesburg - Pretoria.

(3) Where i and j are adjacent regions, they are linked by the shared boundary.

(4) Finally (on the major diagonal), N_{ii} is unique because of autocorrelation effects.

So, the probability of calls arising from any region in the network is *dependent* of the probability of calls terminating in a region of the network. For the four reasons given, the variations between the observed call connections and those predicted can never be explained only in terms of pure chance.

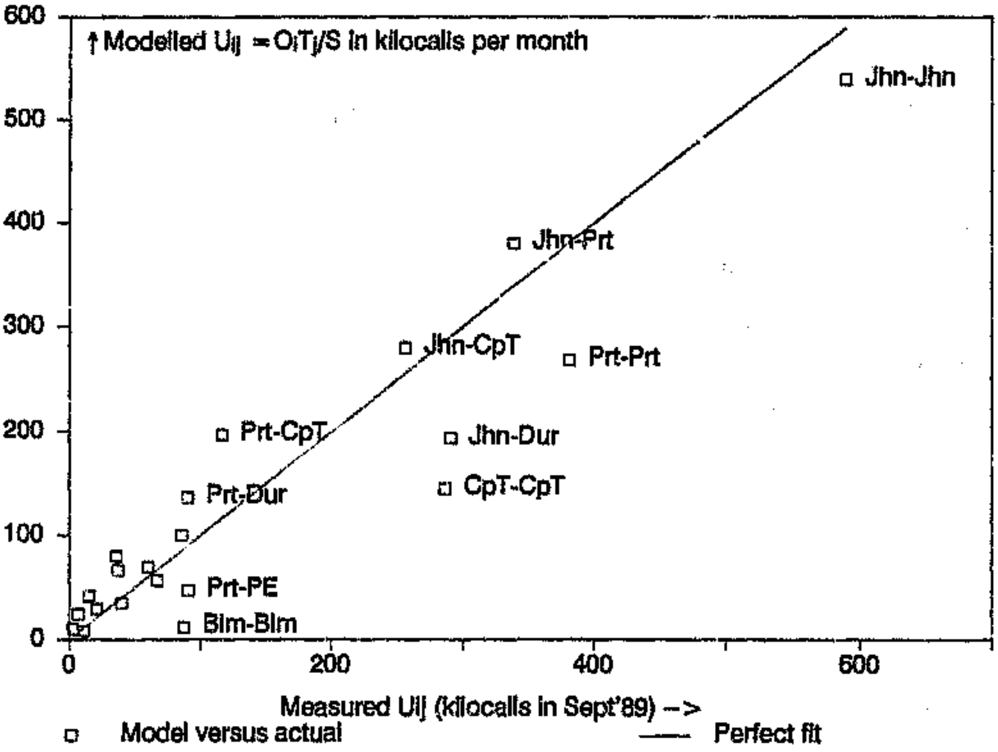

Figure 10. Equation 48 model versus actual U_{ij}

The results of the correlation analysis are plotted in fig. 10. The abbreviations used are explained in table 11. It was found that the value of R^2 was 0,847. To prevent the graph from being cluttered, the *average link usage* has been plotted. This is defined as:

$$'J_{ij} = (N_{ij} + N_{ji})/2 \tag{51}$$

Fig. 10 is interpreted as follows. The horizontal axis shows measured values. The vertical axis shows the expected values. The squares show the scatter of measurements from a model assuming statistical independence. The degree of dependence of origin on destination can be seen from the scatter about the line of perfect fit.

5.6 CONCLUSION

Hypothesis 3 is not supported because the assumption of independence is not satisfied. Yet, the basis of the former Bell System model must not be abandoned. In the real world we need a workable model. We need one which offers the advantages of a known theoretical basis and a simple formulation. As Lee (1973) says, "Essentially, a model is a representation of reality." He continues:

> "It is usually a simplified and generalized statement of what seems to be the most important characteristics of a real-world situation; it is an abstraction from reality which is used to gain conceptual clarity — to reduce the variety and complexity of the real world to a level we can understand and clearly specify."

The assumption of statistical independence yields a useful approximation to reality. In fact it forms an ideal foundation for further study of the four factors enumerated above.

It was found that 84,7% of the sum of squares of variation in the values of N_{ij} is accounted for by an hypothesis of independence of call origin and call destination. This shows that the model is of use, even though the pre-requisite of statistical independence is not met. For example, the call connections from Johannesburg to Johannesburg were 588 940 calls per month. The predicted value is 539 460 calls, being an under-prediction of only 8,4%.

Having disproved independence, it remains to be discovered if distance enters the relationship, and if the use of O_i and T_j can be avoided. These matters are investigated in the next chapter.

CHAPTER 6
ROLE OF RANK, SIZE AND DISTANCE

6.1 REGION RANK, SIZE, INTER-REGIONAL DISTANCE

THIS CHAPTER shows that distance is not a factor in determining call volumes in the packet network. Also the chapter shows that call connections increase with the regions' size or decrease with the regions' rank. Finally the chapter shows that if notice is taken of whether calls are between or within regions, the volume of call connections can be modelled better.

6.2 MULTIPLE REGRESSION MODELS

Multiple linear regression (MLR) is a powerful tool to prove which factors are important. It measures the strength and direction of the effect of each factor. Its successful use relies on a careful, considered choice of factors and form of the model[1]. But it is not a substitute for logical thought. MLR cannot find the best form for a equation: it simply suggests suitable multiplying factors (coefficients) and their signs. MLR assumes an additive model, which may not be the correct form. Thus the use of MLR has been delayed until this chapter.

---

[1] The usual assumption is made that the error terms in the regression model are nearly normally distributed. The statistical distribution of the residuals was not studied.

6.3 HYPOTHESIS FOUR

In all networks encountered in the literature survey, the number of calls decrease with distance. So it seems logical to incorporate inter-regional distance as an explanatory variable.

The fourth hypothesis is that the distance between regions influences call connections in the South African packet network.

This chapter now introduces two novel exogenous variables: the rank of the region in which the calls originate and the rank of the region in which the calls terminate.

6.4 HYPOTHESIS FIVE

It needs to be proved that more calls occur between higher ranking regions and less between lower ranking regions.

The fifth hypothesis is that the number of call connections decrease with the rank of both the originating and terminating regions.

6.5 HYPOTHESIS SIX

Calls between regions seem to differ from calls within a region. Can this be proven?

The sixth hypothesis is that call connections within a region and between regions differ in the number of call connections.

This chapter now introduces a third novel exogenous variable: an inter-regional indicator. The inter-regional indicator is a dummy variable equal to unity if the call was between regions and zero if the call was within a region. Any constant would give the same results, but conventionally a value of unity is adopted.

Model I has the optimal coefficients x_0 to x_6 (given in the first line of Table 13) substituted into equation 52.

6.6 METHOD

Models of inter-regional calls usually have as exogenous variables the logarithms of the originating and terminating totals and the logarithm of the internodal distance. Six models of call connections were tested using multiple linear regression. Exogenous variables x_1 to x_6 were selected from a set of six possibilities: the logarithms of the originating and terminating totals, the inter-regional indicator, the logarithm of the internodal distance, and the ranks of the regions. The class of models tested was of the form:

$$\begin{aligned}\log(N_{ij}) &= x_0 + Ax_1 + Bx_2 + Cx_3 + Dx_4 + Ex_5 + Fx_6 \\ &= x_0 + A\log(O_i) + B\log(T_j) + C\,(\text{indicator}) \\ &\quad + D\log(D_{ij}+1) + E.i + Fj, \qquad (52)\end{aligned}$$

where x_0, A, B, C, D, E and F are coefficients determined from multiple linear regression.

Up to this point, i and j have been regarded merely as being subscripts to the components of the vectors O and T. Now, additional meaning is attributed to i and j. They take on the attribute of rank. This is equivalent to sorting the regions according to size before subscripting. Thus,

i = rank of originating region,
j = rank of terminating region,
O_i = total volume of calls originating in the i^{th} region,
T_j = total volume of calls terminating in the j^{th} region,
and inter-regional indicator = 1 if $i \neq j$ or indicator = 0 if $i = j$.

Logarithms have been taken of variables whose values vary over a wide range. The inter-regional distance D_{ij} was calculated from co-ordinates used for tariff purposes, expressed to the nearest kilometre. See table 12. To avoid taking the logarithm of zero, one kilometre was added to all distances. See equation 52. Model I has the optimal coefficients x_0 to x_6 (given in the first line of table 13) substituted into equation 52.

Table 12. Distance D_{ij} (in km)

	FROM \ TO	Jhn	Prt	CpT	Dur	PE	Blm
Jhn	Johannesburg	0	53	1 264	499	892	370
Prt	Pretoria	53	0	1 311	532	945	422
CpT	Cape Town	1 264	1 311	0	1 280	669	916
Dur	Durban	499	532	1 280	0	689	472
PE	Port Elizabeth	892	945	669	689	0	539
Blm	Bloemfontein	370	422	916	472	539	0

Originating and terminating ranks i and j were temporarily dropped from the generic model I, to form model II. (See missing entries in table 13.) Model II still included the logarithm of the internodal distance.

This function of distance was subsequently dropped to give model III. This caused no loss of accuracy. (See results in table 14.)

Still keeping the inter-regional indicator, logarithms of the originating and terminating totals were replaced by the originating and terminating ranks to form model IV.

Model V used just the logarithms of the originating and terminating totals.

Finally, model VI used ranks but not the inter-regional indicator.

6.7 RESULTS

Tables 11 and 12 appear to have no patterns in common. The coefficients from the multiple linear regression are summarized in Table 13. Table 14 shows the results of the regression.

Table 13. Coefficients for equation 52

		$x_1=$ log(O_i)	$x_2=$ log(T_j)	$x_3=1$ if $i=j$	$x_4=$ log($D_{ij}+1$)	$x_5=$ rank i	$x_6=$ rank j
	Constant x_0	— A	— B	Coefficients of x_1-x_6 C	— D	— E	— F
Model							
I	0,69	0,096 (0,38)	0,52 (0,26)	0,66 (0,18)	0,068 (0,06)	-0,17 (0,09)	-0,09 (0,07)
II	-3,08	0,81 (0,11)	0,91 (0,11)	0,54 (0,18)	0,009 (0,05)		
III	-3,11	0,82 (0,10)	0,91 (0,09)	0,52 (0,14)			
IV	3,13			0,52 (0,14)		-0,20 (0,03)	-0,23 (0,03)
V	-3,03	0,82 (0,11)	0,91 (0,10)				
VI	3,21					-0,20 (0,03)	-0,23 (0,03)

The standard error of each coefficient is given in brackets in smaller type.

Table 14. Results of models using equation 52

Model	R^2 as %
I	82,8
II	80,6
III	80,6
IV	78,5
V	74,3
VI	72,1

6.8 DISCUSSION

The generic model I, with all coefficients, is regarded as being too complicated.

Models II and III yielded the same R^2 (within a decimal of a percent). Eliminating the function of distance caused no loss of accuracy. Thus, hypothesis four cannot be supported. The small value of the coefficient D in model II supports this decision. (See table 13.) This result is surprising as distance was thought to play an important role in every network.

Model IV seems to contain less information than model V, but surprisingly, model IV was more accurate. This result reveals the importance of rank and the inter-regional indicator variable.

The correlation between model V and measurements was less than 75%, the pre-decided level of acceptability. Model VI also did not reach the level of acceptability. Models V and VI are included for completeness, because N_{ii} is undefined in many networks (e.g., in the international network).

6.9 SUMMARY

Hypothesis four is rejected, while hypotheses five and six are supported. Model IV is satisfactory.

6.10 CONCLUSION

In the packet network, inter-regional distance plays no role.

The coefficients A and B are similar, showing near-symmetry in the role of origin and destination in causing call connections.

The negative values for both E and F show that individual ranks must be subtracted (or appear in the denominator) of a model. The size and rank of the destination is only marginally more important than that of the origin in determining call connections. The almost equal values for E and F show that the rank of the origin and the destination can be treated symmetrically in a model that excludes O_i and T_j. The inclusion of an inter-regional indicator keeps the model up to the acceptable level.

The final chapter in this thesis will bring together the hypotheses and create a 2-dimensional rank-size model.

CHAPTER 7
SUMMARY AND CONCLUSIONS

7.1 BRINGING IT ALL TOGETHER

IN THIS CHAPTER, the main results from chapters three to six are brought together.

Six hypotheses support this thesis. Linear regression was used to test the validity of the first three. Multiple linear regression was used to test the validity of the last three.

It has been shown that only hypotheses one, two, five and six are supported. Hypothesis three was rejected by the χ^2 test because the assumption of independence was not satisfied. (Nevertheless, equation 48 gives a useful approximation, explaining 84,7% of the variations.) Also, hypothesis four was rejected as distance proved unimportant.

Hypothesis five was conceived as a 2-dimensional extension of hypothesis two.

Hypothesis six shows that a better explanation results if the major diagonal of the call connection matrix is considered separately.

Thesis: $N_{ij} = \text{Function}(r)$.

Hypotheses:

H1: $U_r \propto GRIP_r$

H2: $GRIP_r = GRIP_1/r$.

H3: $N_{ij} \propto O_i T_j$.

H4: $N_{ij} = F(D_{ij})$.

H5: $N_{ij} = G(1/i, 1/j)$.

H6: $N_{ij} = H(i = j)$.

From chapter three, hypothesis one, equation 41 implies that

$$O_i = K\,GRIP_i \text{ and} \tag{52}$$

$$T_j = K\,GRIP_j. \tag{53}$$

From chapter four, hypothesis two, equation 45 implies that

$$GRIP_i = GRIP_1/i \text{ and} \tag{54}$$

$$GRIP_j = GRIP_1/j. \tag{55}$$

From chapter five equation 48 states:

$$N_{ij} = O_i T_j / S,$$

assuming the approximation of statistical independence of origin and destination. Therefore, by substituting equations 52 and 53,

$$N_{ij} = K\,GRIP_i\,K\,GRIP_j , \tag{56}$$

and by substituting equations 54 and 55,

$$N_{ij} = (K\,GRIP_1)^2/(ijS). \tag{57}$$

For a given network, at a given point in time, S and $GRIP_1$ are constants, yielding:

$$N_{ij} = k/(ij) \tag{58}$$

The value of k can be found by letting i and j be equal to one. Then,

$$N_{11}=k. \tag{59}$$

That is, the dimensioned constant, k, is equal to the volume of calls within the first-ranked region. Therefore the number of calls between two regions is

$$\boxed{N_{ij}=N_{11}/(ij), \qquad (60)}$$

where i and j are the ranks of the originating and terminating regions, and N_{11} is the number of calls within the largest region. Equation 60 is supported by the results of the test of hypothesis five and the similarity of the coefficients E and F in model IV.

Equation 60 yielded an R^2 of 0,761 between model and measurements, which is acceptable. The R^2 values for the hypotheses are summarized in table 15. The table gives the results of linear regression for the first three hypotheses individually and collectively.

Table 15. Summary of strength of hypotheses

Hypothesis	R^2
H1	0,954
H2	0,924
H3	0,847
H1-H3	0,761
H4	Rejected
H5, H6§	0,785

§ The result of the multiple linear regression for model IV is given.

7.2 SUMMARY OF THESIS

Using dimensional, probability and growth analyses, this thesis has found fatal flaws in the textbook gravity model.

The thesis has also measured a network in which distance between regions plays no role in determining the volume of calls between them. This conclusion was based on a complete matrix containing more than four million calls.

Calls between and within regions of a network can be estimated from the economic ranks of the regions. An improvement results if the inter-regional indicator is introduced as a dummy variable.

7.3 THE CONTRIBUTION OF THIS WORK

The unique contribution of this work is that it –

- Chose a new and useful topic for research (economics and its relationship with data calls);
- Named the concepts of unweighted and weighted Gross Regional and Industrial Product, Network Revenue (per industry), Data Network Intensive ratio, and call connections;
- Determined reliable data sources for the above economic and network concepts;
- Created a means to measure call connections and judge whether the measurements are valid;
- Published for the first time a national matrix of calls;
- Showed how data from the sources can be combined usefully and meaningfully to give regional call volumes;
- Showed for the first time, empirically and theoretically, that the best of the existing methods of modelling call volumes between regions is based upon an approximation of statistical independence of origin and destination;
- Modelled without regional sizes by introducing regional ranks;
- Generalized the above results to avoid country-dependent variables and
- By means of equation 60, explained 76,1% of the variations that occur in the volume of calls between and within regions.

7.4 CONCLUSION

It is appropriate now to retrace the path that this thesis has taken. Telecommunication data network planners need a simple model of the network. This thesis has investigated what causes calls between computers and terminals in this country. It has been shown how Gross Regional and Industrial Product ranks (equation 60) influence the resources required for data calls.

First, this thesis studied gravity models. It found that they suffer from serious defects. In particular the gravity models are the application of a three-dimensional equation to a two-dimensional problem. Further, the models fail dimensional analysis if non-integer powers are used. Also the gravity model yields a growth equation that cannot be substantiated from a probability viewpoint.

Second, using dimensional analysis this thesis examined existing growth models. It showed that only one model corresponded to a reasonable probability assumption. The assumption is that calls are generated and terminated independently of each other. Reasons for the failure of the assumption have been stated. However, the assumption of statistical independence is a useful approximation. It makes a good *normative* model.

Third, this thesis has shown that the number of calls to and from a region is proportional to the amount of economic activity in the region. This is measured in terms of Gross Regional Product: the contribution of a region to the Gross Domestic Product (GDP). That is, data calls and economic activity are correlated. Next this thesis showed that the industries contributing to GDP can be given industry dependent weights. These weights improve the correlation between calls and activity because the weighting process takes the profile of the average network user into account.

Fourth, this thesis investigated the mechanism behind the contribution of a region and industry to the GDP, i.e., Gross Regional and Industrial Product (GRIP). The thesis took the population of the centres of South Africa, and showed that it follows the rank-size rule. Next Gross National Product (GNP) for the top countries of the world was shown also to follow the rule. Then unweighted and weighted GRIP for this country were both shown to follow the rank-size rule. So, the rank-size rule may be regarded as a mechanism generating GRIP. This result may be used to generate GRIP algorithmically for simulations, in place of look-up tables.

Fifth, this thesis presented measurements of calls made within and between regions of the South Africa packet network. (This was after a filter was developed to reject invalid call records and the records were grouped into regions.) It was shown that the former Bell system model could be used to supply an approximation to the number of calls.

Sixth, multiple linear regression on the measured call connections showed surprisingly that modelling with inter-regional distance produced no noticeable improvement. Thus this work has discovered a "frictionless" network that can lay the foundations for the study of networks where distance is important. The reason for calls being independent of distance is that the network is not a sociological network, nor is the cost of a call dependent on distance.

Seventh, it has been shown that call connections decrease with the ranks of the regions connected.

In conclusion, this thesis has developed a two-dimensional extension to the rank-size rule. This extension predicts the number of calls on a link between two regions, the ranks of which are known (equation 60).

7.5 RECOMMENDATIONS

As a result of this research it is recommended that—

First: The data sources found and the methods developed in this thesis form the future scientific basis for division of total predicted calls among nodes and links of data networks;

Second: The measured call connections be used as the input to *network performance models*;

Third: A program be set up to filter and report on a daily basis on the type of call records discarded by this study. This will help eliminate maintenance problems and congestion. This should include a study of short and long term fluctuations in calling patterns.

Fourth: Equation 60 be used as a normative model to detect link-specific faults.

7.6 SUGGESTIONS FOR FURTHER RESEARCH

As a result of the research, the following may prove useful avenues for future research.

7.6.1 OTHER DATA

To correctly dimensioning switching equipment, the time profile of calls and the statistical distribution of their lengths are needed. This data can be derived from the source used for this thesis. It will be necessary to use the filter developed in this thesis.

7.6.2 BETTER INSTRUMENTS

There is a need for sharper tools. For example, a system to classify company names according to Standard Industry Code might be useful. It would automate the time-consuming process, and allow such classifications to be debated and replicated.

7.6.3 OTHER NETWORKS

Data can be obtained for other networks and other countries. These can be related to Gross Regional and Industrial Product. The customer profile can be found for voice or other networks (e.g., Metropolitan Area Networks). GRIP can then be weighted using the profile. A ten-industry profile of the customer should improve the modelling. However, for quick modelling, a single industry model (e.g., the financial industry) may give adequate answers. Even if this data is lacking, ranks of regions may be used as surrogate variables. Provided that an inter-regional indicator is used, the model may be just as good.

7.6.4 DISTANCE PROBABILITY

The model can be extended to a *multiple joint probability* equation that invokes the probability of a call traversing a particular distance. This will be of importance, for example, in voice networks. Distances can be bracketed into bands. Then the probability $P(D_{ij})$ that the call will fall in a distance band can be introduced as the last factor in equation 61:

$$P(ij) \approx P(O_i)\, P(T_j)\, P(D_{ij}). \qquad (61)$$

This equation can be rewritten with $P(D_{ij})$ being the subject of the equation and plotted for various values of D_{ij}. A table such as table 10 can be extended from a two-dimensional (origin-termination) table to a three-dimensional table (origin-termination-band). This last equation of the thesis may be regarded as a generalization of the first equation of the thesis. Both equations assume independence of the factors. However, equation 61 generalizes by making no assumptions about the form of the function of distance.

REFERENCES

Abu-Eisheh, S. A. (1987), *Dynamic Traffic Network Equilibrium with Route and Departure Time Choice Model,* Pennsylvania State University, Ph. D. Thesis.

Alesifeer, A. S. (1987), *Spatial Transferability of Trip Generation Models,* University of Colorado at Boulder, Ph. D. Thesis.

Allen, K. M. (1988), Proc. Third International Symposium on Spatial Data Handling, 17-19 Aug. 1988, *Trade Networks and European Contact: A Case Study Using Geographic Information Systems*. Int. Geogr. Union: Columbus, OH, USA, 367-385.

Allison, M. J. (1985), *Forecasting Telephone Traffic—Some Recent Developments,* Telecommunication Journal of Australia. 35:1, 31-39.

Almaani, M. S. S. (1988), *Network Trip Assignment and Aggregation Procedures in Urban Transportation Planning and Design,* University of Pennsylvania, Ph. D. Thesis.

Appelbe, T. W. et al. (1988), *Point-to-Point Modelling: An Application to Canada-Canada and Canada-United States Long Distance Calling,* Conference on Telecommunications Demand Modelling, Information Economics and Policy, 3:4, 311-331.

Armstrong, R. H. (1990), *Network Services,* Tenth Vacation School on Switching and Signalling in Telecommunication Networks. Birmingham, IEE: London 2-7 Sept. 3/1.

Bear, D. (1976), *Principles of Telecommunication Traffic Engineering,* Peter Peregrinus: Stevenage, Herts. England. 180, 186.

Byler, J. and O'Sullivan, P. (1974), *The Forecasting Ability and Temporal Stability of the Coefficients of Gravity Models applied to Truck Traffic,* Traffic Engineering and Control. 10:1, 474-476.

Catanese, A. J. (1972), *Scientific Methods of Urban Analysis,* Urbana: University of Illinois Press, 58.

CCITT, (1968), International Telegraph and Telephone Consultative Committee, *Local Telephone Networks*. July, Chap. 4, Geneva: © International Telecommunications Union. 1-15.

CCITT, (1983), *GAS3: General Network Planning,* Geneva: © International Telecommunications Union, 232-235.

CCITT, (1988), *GAS9: Case study on the Economic and Technical Aspects of the Transition of a Complete Analogue National Network moving to a Digital Network,* Geneva: International Telecommunications Union, 40.

Chiu, C-H. (1987), *The Development and Calibration of a Peak-Period Work Trip Distribution Model: An Empirical Case Study,* University of Pennsylvania, Ph. D. Thesis.

Christian, W. and Braden, W. (1966), *Rural Migration and the Gravity Model,* Rural Sociology. 31:March, 73-80.

Combot, J-P., Debois, J-P. and Treps, D. (1988), *SCOOP – Switching Centers Operating Observations Processing,* IEEE J. on Selected Areas in Communications. 6:8, 1371-1377.

Congdon, P. (1989), *Modelling Migration Flows between Areas,* Regional Studies. 23:2, 87-103.

CSS (1985), Central Statistical Service, *Population Census, Report No. 02-85-02 Age by Development Region, Statistical Region and District,* Private Bag X44, Pretoria.

CSS (1988), *SIC: Standard Classification of all Economic Activities,* Pretoria. 4th Ed., Oct.

CSS (1989a), Central Statistical Service, *User's Guide,* Pretoria, December.

CSS (1989b), *Gross Geographic Product at Factor Incomes by Magisterial District - 1984,* Statistical News Release P0401, Pretoria, 10 July.

CSS (1989c), *Standard Code List for Statistical Regions, Magisterial/Census Districts, Cities, Towns and Non-urban Areas,* 9th Ed., Pretoria, 3 January.

Curry, L. (1972), *A Spatial Analysis of Gravity Flows,* Regional Studies. 6:131-147.

David, A. J. and Pack, C. D. (1979), *The Sequential Projection Algorithm: A New and Improved Traffic Forecasting Procedure,* National Organizing Committee, 9th Teletraffic Congress. 8 pp.

Davis, J. T. (1965), *The Scientific Approach,* London: Academic Press. 60.

Defris, L. V., Layton, A. P. and Zehnwirth, B. (1986), *The Impact of Economic Cycles on the Demand for International Telecommunications in Australia,* Information Economics and Policy. Elsevier Science B.V. North Holland. 2:2, 105-117.

Debbiesse, J. L. and Matignon, G. (1979), *Comparison of Different Methods for the Calculation of Traffic Matrices,* National Organizing Committee, Ninth Teletraffic Congress. 9 pp.

Drew, D. R. and Chen, W-C. (1976), *System Dynamics Simulation as a Methodology for Transportation Systems Analysis,* Modeling and Simulation. 7:411-416.

Dunstan, A. W. (1977), *A Mathematical Model of Telephone Traffic Dispersion in some Australian Metropolitan Networks,* Australian Telecommunication Research. 11:1, 84-91.

Fotheringham, A. S. (1981), *Spatial Structure and Distance-Decay Parameters,* Annals of the Association of American Geographers 71:Sep. 425-436.

Fotheringham, A. S. (1984), *Spatial Flows and Spatial Patterns,* Environment and Planning A. 16:529-543.

Furness, K. P. (1965), *Time Function Iteration,* Traffic Engineering and Control. No. 7, 458-460, Nov.

Hoover, E. M. and Giarratani, F. (1985), *An Introduction to Regional Economics,* Third ed., New York: Knopf. 211.

Housner, G. W. and Hudson, D. E. (1959), *Applied Mechanics. Dynamics.,* Princeton, NJ. Van Nostrand. 2:5-25.

Hua, C-i. and Porell, F. (1979), *A Critical Review of the Development of the Gravity Model,* International Regional Science Review. 4:2, 97-126.

Jacobaeus, C. (1980), *Teletraffic Theory and its Practical Applications,* Ericsson Review. No. 1, 8-15.

Jipp, A. (1960), *Estimating the Volume of Traffic by a Simple Formula,* Telecommunication Journal. 8e-11e. Jan.

Kennedy, I. G. (1991a), *NetCalc,* A program to aid the network planning process. Version 1.07.

Kennedy, I. G. (1991b), *Division of Labour and Allocation of Resources for Network Management,* Joint SAIEE/CSSA International Symposium on Network Management. 27-29 May. 6.1-6.15.

Kennedy, I. G. (1991c), *Weighted Gross Geographic Product as a predictor of X.25 customer calls,* Trans. SAIEE. 82:2, 123-128.

Kennedy, I.G. (1992), *Economic, Regional and Industrial Factors Affecting Calls,* Third Africon Conference. New York: IEEE 92 CH 3215-1. 307-310.

Kittel, C., Knight, W. D. and Ruderman, M. A. (1965), *Mechanics,* Berkeley Physics Course, Vol. 1. New York: McGraw-Hill, 264.

Knox, P. L. (1980), *Measures of Accessibility as Social Indicators: A Note,* Social Indicators Research. 7: Jan. 367-377.

Kruithof, J. (1937), *Telefoonverkeersrekening,* De Ingenieur. (In Flemish.) 52: Feb. 8, E15 - E25.

Kurian, G. T. (1979), *The Book of World Rankings,* London: Macmillan Reference Books. 81.

Kvalseth, T. O. (1985), *Cautionary note about R squared,* American Statistical Association, 39:4, part 1, 281.

Lee, C. (1973), *Models in Planning,* Urban and Regional Planning Series Vol. 4. Oxford: Pergamon, 7.

Leik, R. K. and Meeker, B. F. (1975), *Mathematical Sociology,* Englewood Cliffs, NJ: Prentice-Hall, 15.

Levenbach, H. and Cleary, J. P. (1984), *The Modern Forecaster: the Forecasting Process through Data Analysis,* New York: Van Nostrand Reinhold.

Liang, X. J. (1988), *Method for Local Network Planning,* Torino, June 1-8, 1988. Organizing Committee, Twelfth Teletraffic Congress. (Twelfth International Teletraffic Congress. Proc. 3.2B.3) B3.1 - B3.4.

Mathur, M. C. and Satsangi, P. S. (1985), *Entropy-Maximizing Methods of Transport Analysis vis-a-vis a Physical System Theory Modeling Framework,* IEEE Transactions on Systems, Man, and Cybernetics. SMC15:2, Mar-April, 281-290.

Mohr, P. J. *et al.* (1988), *The practical guide to South African Economic Indicators,* Johannesburg: Lexicon, 40-41.

Mueser, D. (1989), *The Spatial Structure of Migration: An Analysis of Flows Between States in the USA over Three Decades,* Regional Studies. 23.3, 185-200.

Neal, F. and Shone, R. (1976), *Economic Model Building,* London: MacMillan. 165.

Newstead, I. A. (1961), *Review of Telephone Traffic Engineering — Part II,* Telecommunication Journal of Australia. 13:1, June, 38-48.

O'Neill, W. A. (1987), *Origin-Destination Trip Table Estimation Using Traffic Counts,* State University of New York at Buffalo, Ph. D. Thesis.

Oppenheim, N. (1980), *Applied Models in Urban and Regional Analysis,* Englewood Cliffs, NJ: Prentice-Hall.

Orvis, W. J. (1987), *1-2-3 For Scientists and Engineers*, San Francisco: Sybex, 164.

Ostle, B. and Mensing, R.W. (1975), *Statistics in Research*, Ames: Iowa State University, 540-541.

Palmowska, K. (1981), *A Study of Traffic Distribution Models with Reference to the Polish Telex Network*, Rozprawy Elektrotechniczne. (In English.) 27:2, 593-605.

Purcell, E. M. (1965), *Electricity and Magnetism*, Berkeley Physics Course, Vol. 2. New York: McGraw-Hill, 27.

Rahi, M. Y. (1987), *Estimation of Origin-Destination Trip Matrix from Link Traffic Information*, Washington State University, Ph. D. Thesis

Rapp, Y. (1962), *Calculation of Traffic Distributions in Multi-Exchange Networks.*, Stockholm. Ericsson Technics. 18:1, 4-33.

Reason, J. (1987), *The Psychology of Mistakes: A Brief Review of Planning Failures*, Chapter 5 in Rasmussen, J., Duncan, K. and Leplat, J., Ed. *New Technology and Human Error*, Chichester: Wiley, 45-52.

Rich, D. C. (1980), *Potential Models in Human Geography*, Concepts and Techniques in Modern Geography, No. 26. 19.

Saaty, T. L. and Alexander, J. M. (1981), *Thinking with Models: Mathematical Models in the Physical, Biological, and Social Sciences*, Oxford: Pergamon Press. 150.

Saito, H. (1986), *An Adaptive Traffic Forecasting Method with Macro-Traffic Forecasts*, Trans. Inst. Electron. & Commun. Engineers of Japan. Part B. (In Japanese). J69B:2, Feb, 123-129.

SAPT. (1991), *Telematic Services AGM's Report*, 13, 17 April.

Scheaffer, R. L. and McClave, J. T. (1982), *Statistics for Engineers*, Boston: Duxbury Press. 228.

Smith, S. L. J. (1985), *US Vacation Travel Patterns: Correlates of Distance Decay and the Willingness to Travel*, Leisure Sciences. 7:2, 151-174.

Smith, T. S. (1976), *Inverse Distance Variations for the Flow of Crime in Urban Areas*, Social Forces. 54:4, June, 802-815.

Stewart, J. Q. (1947), *Empirical Mathematical Rules concerning the Distribution and Equilibrium of Population*, Geographical Review. 37:461-485.

Taylor, P. J. (1971), *Distance Transformation and Distance Decay Functions*, Geographical Analysis. 3:221-238.

Terranova D. and Ulian, G. (1989), *A Model to Forecast Telecommunication Trends in Scarcely Defined Contexts*, Telettra Review, 44:1, 3-16.

Tocalis, T. R. (1978), *Changing Theoretical Foundations of the Gravity Concept of Human Interaction*, Northern Illinois University Press. (In Berry, J. L. Ed. *The nature of Change in Geographical Ideas*, DeKalb.) 123.

Visick, G. (1980), *A Modification to Kruithof's Double-Factor Method*, Transportation Research Part B. 14B:4, 307-318.

Wheelwright S. C. and Makridakis, S. (1985), *Forecasting Methods for Management — 4th Edition*, New York: Wiley.

Whitmore, R. E. (1965), *Graphical and Mathematical Investigation of the Differences in Traveltime Factors for the Gravity Model Trip Distribution Formula in Several Specific Urban Areas*, Knoxville. Tenessee University. Apr. (Available from NTIS as PB 182 445.) 127 pp.

Wood, W. G. and Martin, D. G. (1974), *Experimental Method*, London: Athlone Press, University of London.

APPENDIX

Gross Regional and Industrial Product (GRIP) in kilorands p.a. 1984 from CSS (1989b)

- See table 2 on page 48 for meaning of abbreviations of industries.
- The imputations column should be subtracted in calculating Gross Regional Product (GRP).
- The totals for the regions give unweighted GRIP (GRP) for the CSS region.
- Negative GRIP indicates a nett loss in the region by the industry during the year.
- The CSS regions are ordered according to CSS region number (not shown).
- The regions are grouped according to CSS statistical regions.

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Bellville	1604871	55021	8058	513491	38534	68673	189888	93698	327798	31138	56292	281755	53109
Goodwood	468924	0	0	192144	779	17064	116771	11913	59029	8314	10137	47871	25176
Cape Town	5506187	62914	2865	1044342	171265	185617	988208	1045199	1262348	102509	216779	690969	166710
Simons Town	200423	1017	0	6380	360	7593	13736	4099	19992	1995	3433	139346	8738
Wynberg	1015204	11879	3709	290096	0	88614	187936	42099	205615	33155	353[illegible]	146039	41372
Kuils River	170192	16806	4149	68554	579	9030	19928	10784	13184	1676	2264	22564	5202
Paarl	656131	239909	99	167068	5086	22153	74643	9770	81773	5132	14043	56742	7799
Somerset West	194135	16787	1807	82450	2421	7418	20331	6856	21399	4477	3675	28081	5783
Stellenbosch	455431	92910	0	57683	3287	11723	74977	9805	59002	4221	10132	71689	80266
Strand	98163	2041	0	23480	1611	6306	19074	3487	27958	1912	4801	12099	4996
Wellington	73943	5916	0	28587	1799	1319	7844	7995	9446	805	1622	9546	2308
Bredasdorp	80381	54968	1125	428	442	1971	6512	1806	6523	777	1120	5164	1485
Caledon	202872	103995	35	22922	2214	3499	23100	8298	17198	1871	2953	18486	4207
Heidelberg	28623	15501	998	231	331	305	2918	1388	5067	99	870	1800	855
Hermanus	38024	7396	0	1713	1157	2245	7421	2043	8559	1200	1470	6022	1738
Swellendam	93119	41495	0	13809	642	1487	8845	3782	11461	729	1968	10541	2296

Appendix

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
George	207971	24044	1070	25966	5113	15275	29016	8560	47410	4302	8141	49408	5948
Knysna	92199	8009	1475	14889	2892	9959	15595	4709	14134	1074	2427	19122	2768
Mossel Bay	123294	17347	19485	18590	2695	5009	16215	15007	14309	798	2457	13923	2373
Riversdale	51003	17079	351	1301	490	1356	6407	3677	7817	1548	1342	9944	2375
Uniondale	15961	8870	0	501	0	63	1928	1359	1676	142	288	1205	505
Calitzdorp	10434	6323	0	417	174	80	439	214	1385	130	238	920	590
Ladismith	31079	14847	0	3867	139	2649	2474	1387	3226	251	554	1985	808
Oudtshoorn	162705	22546	68	2417	703	4863	23814	16723	24134	2491	4144	40812	4471
Ceres	114262	70114	9	11747	1233	1637	8949	3294	7757	681	1332	8284	1869
Montagu	35508	9787	0	9038	768	391	3609	2122	5237	542	899	3682	1231
Robertson	74014	36711	3386	4561	1250	1930	7426	2423	7808	817	1341	6928	2115
Tulbagh	55613	17426	0	19558	367	720	4792	1853	5316	323	913	4649	1522
Worcester	406694	160385	1746	32339	6550	16714	37445	55639	41220	3647	7079	46840	11248

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Input	Gover	Produc
Hopefield	38779	13086	0	160	1289	469	2473	619	2666	36	458	17720	719
Malmesbury	481783	153573	880	143564	103386	9563	24445	9659	15843	3008	2721	15698	4885
Piketberg	96974	32739	2063	29199	837	1047	8285	5661	9825	1046	1687	5187	2772
Vredenburg	150835	27145	1208	60775	505	3551	13765	7031	21469	950	3687	16071	2052
Clanwilliam	89461	53321	0	14066	448	398	6752	2328	5980	632	1027	4732	1831
Vanrhynsdorp	30430	9888	1695	1028	122	142	2771	7758	2310	520	397	3707	886
Vredendal	70582	27110	5810	3887	420	2609	8627	2703	8848	1439	1519	8753	1895
Namakwaland	198431	43280	144119	2079	348	2525	17930	12254	9108	1671	1564	10076	3165
Walvis Bay	108867	7617	2934	15144	2886	1817	20688	23758	16882	433	2899	17211	2386
Beaufort West	80350	2741	0	1533	1132	1310	9520	36475	14598	763	2507	12802	1983
Fraserburg	5065	1360	0	5	129	72	817	464	1489	200	256	1246	539
Laingsburg	10739	1433	0	47	73	434	1320	4029	2067	326	355	904	461
Murraysburg	9229	6261	0	125	50	0	487	232	1032	138	177	632	449
Prince Albert	11657	5457	0	113	108	165	890	810	2059	186	354	1389	834
Victoria West	19674	5542	0	312	198	206	2988	4533	3364	182	578	2391	736

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Input	Gover	Produc
Calvinia	18622	-5818	797	247	328	329	5730	2698	8168	564	1403	5257	1725
Sutherland	4460	467	0	9	40	239	809	307	1331	97	229	944	446
Williston	1040	-3888	0	6	53	169	1136	922	1615	95	277	674	535
Carnarvon	10816	-506	1026	252	122	584	2514	1205	3269	170	561	1661	1080
Prieska	77081	-913	57016	341	518	130	5108	3568	4477	202	769	5980	1423
Britstown	7883	2112	273	102	72	54	1020	899	1900	268	326	915	594
Colesburg	30498	12950	0	129	4795	325	2727	2340	3075	304	528	3067	714
De Aar	80494	3882	754	1703	1046	1152	7958	42652	10342	818	1776	10610	1353
Hanover	6931	2073	0	0	34	2824	302	234	819	4	141	501	281
Hopetown	11757	3620	250	207	114	33	1502	2072	1970	89	338	1505	733
Noupoort	27100	1203	0	68	157	473	1745	18416	2731	174	469	2149	453
Philipstown	19458	3029	0	390	5929	17	637	892	1314	41	226	6928	507
Richmond	10919	4724	0	247	42	574	852	1112	1656	220	284	1146	630
Gordonia	180744	11776	1175	19297	3569	9103	32505	11625	42842	2837	7357	48062	6310
Kenhardt	71175	-1839	62734	174	194	92	1883	2091	2387	140	410	2610	1119

Appendix

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Hay	17971	11090	187	4	84	172	1502	570	1590	57	273	1945	1043
Barkly West	51455	6637	13716	16567	1155	220	2811	3287	2910	298	500	3258	1096
Hartswater	75650	24458	0	3427	2149	314	16396	4152	6807	765	1169	16267	2084
Herbert	19661	6304	303	2467	1280	72	1587	1269	2778	214	477	2738	1126
Warrenton	35483	10260	963	1516	400	243	3591	10008	3817	343	656	3439	1559
Kimberley	566693	13693	16766	39094	25445	13809	106116	145106	80206	7685	13774	120057	12490
Boshof	13214	-1370	1354	212	144	2790	2223	990	2914	235	500	2477	1745
Jacobsdal	5547	455	84	2	9	110	361	723	1030	86	177	2309	555
Koffiefontein	9212	3636	0	162	185	29	920	729	1931	211	332	1249	492
Kuruman	106366	-8513	58256	2446	847	1847	20988	6708	12941	476	2222	10229	2363
Postmasburg	337553	18581	252882	2246	3064	2077	10968	20684	8016	959	1377	15160	4293
Vryburg	101836	6095	16814	7608	2414	3200	22962	13544	17981	1925	3088	8279	4102

Appendix

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Input	Gover	Produc
Bloemhof	18802	1408	501	3347	394	1272	2637	1937	3129	1088	537	2952	674
Christiana	27436	8615	153	1579	300	252	4173	2393	3860	187	663	5798	789
Coligny	11296	-5719	3	9	400	0	2029	8340	2676	245	460	2697	976
Delareyville	14713	-11222	71	481	486	665	8258	3241	5469	910	939	5483	1810
Koster	10732	-4014	297	866	313	124	4465	2147	3219	163	553	2414	1291
Lichtenburg	154021	-27849	3317	92718	1982	4806	3[illegible]12	4143	23383	1348	4016	16316	4061
Potchefstroom	515682	12487	89199	83130	7864	13925	49162	18103	58618	6488	10066	139[illegible]17	47555
Schweizer-Reneke	719	-24708	219	1468	279	280	7460	3522	5317	510	913	5814	1471
Ventersdorp	18810	-2701	1024	351	494	533	7368	1915	4360	315	749	4493	1407
Wolmaransstad	39697	7145	824	1017	698	558	7993	2967	4881	537	838	11549	2366
Marico	112849	41531	9287	19561	1204	629	8436	8962	8228	864	1413	13213	2[illegible]47
Rustenburg	613057	19098	279069	64584	5937	31601	70879	27915	56993	7403	9787	52367	6998
Swartruggens	17564	328	1804	954	249	2576	2021	1546	4183	155	718	3696	770
Klerksdorp	2815350	2[illegible]722345395		53564	31368	41144	102575	31661	113434	13296	19480	86565	13656
Odendaalsrus	149705	14667	75261	6433	1292	3492	14297	10258	11329	1189	1945	10002	3430
Virginia	654991	897	586238	6995	2015	5105	16260	5673	17981	1495	3088	12293	3127
Welkom	1966728	-124681542733		77471	10882	79295	115073	24832	77824	6128	13365	51810	8513

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Bothaville	68745	25810	12	5573	564	1714	16215	1798	11806	850	2027	4632	1798
Bultfontein	33303	15898	0	452	220	67	5898	1158	2956	541	508	5461	1160
Heilbron	70215	31579	1039	14351	474	637	6754	2114	5344	837	918	6192	1812
Hennenman	25682	-1937	531	11606	426	705	4215	3603	4345	418	746	1544	972
Hoopstad	27196	-5724	0	230	125	400	4299	887	2561	241	440	23842	775
Koppies	26696	12429	386	2772	285	1	3379	2316	2905	236	499	1631	855
Kroonstad	293474	9079	1485	27183	3554	4790	45638	68645	60402	4746	10373	72158	6167
Parys	66986	20061	0	13132	1798	1681	9351	2561	11668	1100	2004	5649	1989
Theunissen	42146	27069	137	2402	305	254	3459	2134	2689	273	462	2758	1128
Ventersburg	3369	-3506	0	169	94	86	1632	511	1937	7	333	1927	845
Viljoenskroon	98986	31240	9482	8233	32497	560	6657	2699	3952	489	679	2137	1719
Vredefort	11382	3466	0	1608	144	6	2322	1177	1195	5	205	1083	581
Wesselsbron	41496	24216	0	4058	279	446	5956	1080	2859	243	491	1891	959

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Bethlehem	256006	60658	653	25695	3144	4149	39718	42548	41285	3236	7090	36917	5093
Fiksburg	60306	16979	0	1584	539	1093	14609	3200	17095	609	2936	6253	1281
Fouriesburg	21823	16728	0	381	51	0	775	1144	1085	101	186	1332	412
Frankfort	38170	11010	53	3570	501	963	6064	3182	6869	419	1180	4614	2135
Harrismith	165454	73784	425	35483	1428	5191	14656	6576	11193	978	1922	14461	3201
Lindley	49547	28068	0	809	275	99	4272	4865	5273	610	905	4421	1760
Reitz	23179	225	0	647	348	484	6371	2385	5921	702	1017	5618	1495
Senekal	87384	51011	0	11814	411	697	9566	1955	5510	824	946	4371	2171
Vrede	69748	48324	0	270	275	359	7765	1319	5536	402	951	4519	1930

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Botshabelo	6459	634	0	565	0	0	71	119	3385	17	581	2137	112
Brandfort	19946	5574	518	94	189	223	2269	2826	3243	456	557	3801	1310
Clocolan	35486	24149	0	104	187	11	3888	1262	2586	462	444	2463	818
Dewetsdorp	20667	9325	0	8	70	308	1539	772	1508	127	259	6456	813
Edenburg	11055	6592	0	9	62	4	557	338	1836	56	315	1384	532
Excelsior	31014	16686	0	599	88	237	3988	1635	3776	266	648	3276	1111
Jagersfontein	8401	3180	0	112	64	10	704	927	1505	99	258	1606	452
Ladybrand	48874	25276	541	2156	450	230	3736	2982	4679	363	803	7477	1787
Marquard	4478	-3440	0	375	135	121	2329	603	2785	120	478	1149	779
Petrusburg	13459	5875	71	208	94	83	1490	954	1681	21	289	2484	787
Phillippolis	10171	3592	0	313	7	0	701	734	1148	72	197	3438	363
Reddersburg	6491	2675	0	60	62	89	630	293	714	107	123	1384	600
Trompsburg	7796	3442	0	280	60	47	410	582	1290	112	221	1211	583
Wepener	18248	6461	0	219	132	45	2932	1285	2231	110	383	4426	790
Winberg	29847	16201	0	1644	225	47	2521	912	3053	282	524	4191	1295
Bloemfontein	1496218	40457	2930	94848	19264	43961	225734	302961	287725	32043	49410	418185	77520
Fauresmith	6043	552	13	28	78	0	726	729	1263	203	217	2009	659
Smithfield	14947	9878	0	16	63	6	849	625	2091	194	359	1113	471

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Bethulie	12045	-254	0	143	80	92	1076	2624	3111	127	534	4966	614
Rouxville	29326	18753	0	39	32	86	800	709	1790	224	307	6561	639
Zastron	30680	3046	0	1845	149	47	1939	2820	2505	341	430	2740	678
Albert	26691	4981	231	1146	372	729	3986	5021	4322	496	742	4945	1204
Aliwal North	51101	11698	0	3912	498	530	7093	3238	6868	1942	1179	15110	1391
Lady Grey	9136	4529	0	909	53	14	879	342	851	110	146	854	741
Steynsburg	13130	3676	0	0	98	84	1154	684	2765	35	475	4538	571
Venterstad	5282	1682	0	33	55	0	530	197	1031	11	177	1635	285
Barkly East	24320	12609	0	133	66	0	1415	648	3353	213	576	5721	738
Elliot	20919	11297	0	150	156	29	1578	1027	4650	169	627	2666	824
Indwe	8059	3599	0	2	55	0	758	538	996	158	171	1644	480
Maclear	13908	3076	0	299	49	93	2705	1561	1813	246	311	3184	1193
Wodehouse	17447	9075	0	236	97	1475	1277	855	1883	259	323	1502	1111

Appendix

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Cathcart	16501	7164	1	230	122	153	1775	912	2145	160	368	3395	812
Komga	16127	8578	2	618	86	7	1545	849	2058	25	353	1983	729
Molteno	13183	6473	0	1970	75	48	1086	628	1207	149	207	1089	665
Queenstown	166494	9206	350	976	3891	6020	30607	36366	27048	2293	4645	42079	3496
Sterkstroom	12205	4939	0	35	54	90	698	1249	4748	80	815	734	393
Stutterheim	27691	5517	0	4872	178	287	3055	2023	4259	247	731	6707	1278
Tarka	11352	6274	0	0	52	5	1229	396	1496	160	257	1200	797
K.William's Town	130130	7290	888	22390	2621	2143	23678	6372	24074	10292	4134	31531	2985
East London	1147859	14251	2835	339715	30669	38563	156038	235708	173307	15637	29761	147856	23041
Adelaide	19232	9299	0	76	130	110	2384	1199	2759	212	474	2617	920
Albany	158156	18454	139	6736	3505	6368	17223	8084	25275	3010	4340	49546	24156
Alexandria	45345	34370	0	456	181	800	2521	1450	2212	385	380	2211	1139
Bathurst	22757	5050	202	102	997	2224	3797	1239	3334	469	572	4397	1518
Bedford	12070	3870	0	168	65	380	925	548	1938	1495	333	2392	622
Fort Beaufort	35133	5054	0	799	345	1706	3336	1560	3637	508	625	17967	846
Hofmeyer	3650	706	300	335	25	0	523	228	742	4	127	626	288

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Credock	79106	16703	0	930	854	857	[illegible]	21724	18214	874	3128	12747	2027
Middelburg	41164	4003	0	4833	292	691	4508	4283	4877	881	837	16474	1159
Somerset East	48950	16939	0	1440	624	1794	4864	7791	5889	683	1011	8311	1626
Kirkwood	44444	27210	0	2077	721	170	2788	940	2767	312	475	6492	1442
Aberdeen	17256	11949	0	11	19	51	1114	378	2360	106	405	1030	643
Graaff Reinet	64072	17953	0	1669	798	1100	9972	5299	12654	721	2173	14016	2063
Pearston	7817	5509	0	11	34	49	309	204	921	131	158	493	314
Jansenville	22306	8820	0	14	38	152	1490	8077	1410	289	242	1426	832
Steytlerville	9759	6852	0	0	31	21	565	345	836	72	144	767	414
Willowmore	14313	7910	0	145	118	0	1496	1229	2029	258	348	763	713
Hankey	59415	42616	881	2011	643	320	2932	774	1432	227	246	6790	1035
Humansdorp	71016	12029	0	11411	1002	3189	10425	8888	5774	1827	991	15057	2405
Joubertina	40503	26654	0	4458	41	91	3957	728	2125	301	365	1617	896

Appendix

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Port Elizabeth	2579971	20464	7159	848986	53682	93348	355934	414995	392326	46216	67373	544179	70055
Uitenhage	493140	12622	174	248158	7097	7453	36052	67749	55082	8184	94[illegible]	48842	11186
Durban	6530355	4336	12427	2186520	90623	291663	953786	1202963	913613	142319	156891	688065	200931
Inanda	236993	20819	557	113804	1732	7603	29796	10770	31965	5442	5489	13408	6586
Pinetown	939533	-3686	990	452980	18050	73239	136617	26495	163278	9656	280[illegible]	66464	23489
Lower Tugela	197427	33974	520	76837	4003	7494	25331	16845	11202	3840	1924	15349	3956
Camperdown	159863	29649	427	91708	5785	3526	6303	6688	7379	701	1267	6342	2622
Pietermaritzburg	1222879	25887	1501	310780	22944	40592	176794	153087	203437	22477	34935	263408	36907
Umzinto	200254	36171	235	92647	1618	6193	18159	10145	14366	1995	2467	17266	3926
Alfred	16739	3257	0	5501	280	445	2864	935	1170	369	201	1238	881
Port Shepstone	226509	26374	9867	37095	2613	16456	36896	18388	44074	3701	7569	31306	7308
Mount Currie	80335	31432	164	3946	823	873	14507	7235	6250	1237	1073	12405	2536

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Polela	5606	1368	0	0	0	0	1066	944	495	129	85	1025	664
Underberg	16377	8080	0	84	8	66	3265	1038	1115	160	191	2314	438
Ixopo	48427	27185	0	1903	274	842	5693	3299	3436	616	590	4355	1414
Kranskop	8736	2976	0	707	16	568	1120	1121	532	178	91	1234	375
Lions River	92830	41982	0	15237	1355	1779	4392	4235	6766	929	1162	14573	2744
New Hanover	76708	29714	187	25047	545	2436	5655	4528	2980	390	512	4248	1490
Richmond	34601	21979	143	1931	596	1265	2578	728	1797	235	309	2570	1088
Umvoti	86026	22024	0	28111	627	1188	8275	4494	7422	1327	1274	12259	1573
Mooi River	60513	32624	0	15328	275	90	2704	2274	3498	203	601	3221	897
Impendle	12570	10743	0	156	3	0	835	82	277	3	48	312	205
Bergville	73420	20933	0	0	35140	38	4598	1702	1773	132	305	8638	771
Estcourt	177604	26874	386	52980	29493	3900	15739	8900	12397	1167	2129	22866	5031
Klip River	276753	19519	763	72890	4155	11313	26302	72227	31269	2525	5370	38444	2716
Weenen	12954	8315	0	286	14	0	942	227	1105	433	190	1528	294

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Input	Gover	Produc
Dannhauser	82871	12348	58883	1593	82	57	4208	1847	1413	379	243	1624	680
Dundee	115602	23844	7364	19509	1799	2405	14650	13778	12635	1289	2170	18790	1709
Glencoe	51458	3071	-85	8966	433	932	6184	21647	4537	323	779	5284	945
Newcastle	488146	24967	23598	200508	56570	14356	45794	19780	66510	4065	[illegible]	36585	6835
Utrecht	87791	35550	30737	121	106	296	2475	969	3557	70	611	13604	917
Babanango	14037	11700	0	0	0	1352	300	35	239	0	41	252	200
Ngotshe	44038	21513	10891	762	0	0	749	138	720	9	124	8765	615
Paulpietersburg	35194	19520	1011	1814	63	4779	2929	785	1890	345	325	1806	577
Vryheid	353717	40978	135629	23874	143	17995	28266	59249	21387	2397	1673	23470	4002
Eshowe	83613	13807	0	26314	550	1467	9873	3215	6961	475	1195	20394	1732
Hlabisa	39710	10330	0	2783	332	926	11422	2611	4894	519	841	5453	1281
Lower Umfolozi	554911	35821	42335	186107	9475	36930	41574	132166	36452	5923	6260	28783	5611
Mtonjaneni	21397	11837	0	390	190	169	3240	837	1656	621	284	1368	1373
Mtunzini	128536	25942	18	77825	446	5619	5726	4821	3868	801	664	2159	1975

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Amersfoort	26430	17016	0	459	0	1080	2068	711	1182	1191	203	2417	509
Bethal	1029160	93757	138283	11873	695923	5582	25921	12915	23294	2837	4000	16956	5819
Carolina	61151	35191	-24	3412	443	1311	5229	3090	4072	734	699	7357	1135
Ermelo	454663	91319	63030	20772	143610	8181	25246	26921	39913	3442	6854	34631	4452
Piet Retief	202731	63668	8015	55565	1120	13870	15859	8314	11500	1611	1975	23042	2142
Standerton	245400	89156	8821	35319	2555	35695	19323	10326	18115	1802	3111	23211	2988
Volksrust	44841	7762	0	875	623	1747	7835	11110	5496	655	944	8268	1414
Wakkerstroom	17727	13094	0	79	46	398	806	650	534	259	92	1508	445
Balfour	182312	20065	16994	7234	117232	665	6132	3048	5304	445	911	4363	1741
Highveld Ridge	2332044	5706	651730	1566970	5110	18395	13197	28825	7896	4262	1356	23005	8304
Belfast	74627	32858	5260	7277	580	416	6867	5614	6271	524	1077	8857	1180
Groblersdal	127204	49575	2291	7824	821	3036	17908	5927	10332	2555	1774	26966	1743
Middelburg	1202161	101715	310316	98453	519900	13565	39486	18754	41766	9337	7172	48675	7366
Moutse	10803	6760	0	39	0	3	452	119	1447	3	248	1814	414
Waterval Boven	23886	5217	0	211	139	42	2592	10641	2888	131	496	2065	456
Witbank	1441974	53597	389924	306125	421538	53301	59900	44063	61355	8599	10536	46955	7152

Appendix

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
Barberton	254478	71201	82890	38078	1495	5835	18027	24909	10176	1291	1747	29639	2684
Lydenburg	226440	33069	58762	73621	1582	1648	17266	9839	12855	1371	2208	16271	2364
Nelspruit	355717	38706	3018	92482	3028	21531	57758	52222	41989	6445	7211	42346	3403
Pilgrim's Rest	122971	19899	337	23901	425	2212	13200	8718	5780	509	993	47154	1829
Witrivier	115697	37547	78	18215	911	2228	18795	4788	5696	962	978	25249	2206
Letaba	26[illegible]	85705	757	46816	3922	9862	43905	15902	27307	2621	4689	24535	3507
Messina	[illegible]	-4192	8932	537	179	322	5030	4012	5992	443	1029	4874	877
Phalaborwa	385322	5837	292371	30170	5066	2642	6281	2852	10407	2279	1787	27414	1790
Pietersburg	453590	31991	28179	50843	4303	31865	73401	55629	74765	17697	12839	91335	6421
Soutpansburg	116895	12280	47	18435	2970	2454	23964	12863	17278	1363	2967	25502	2706
Potgietersrus	188157	35256	28639	19241	3262	5605	34503	14325	21117	2138	3626	23110	4587
Waterberg	155814	8055	54405	2816	13125	26832	12290	8657	14237	1117	2445	13926	2799
Thabazimbi	43634	-24337	37730	1149	1046	2907	4729	2765	6255	948	1074	9033	2433
Warmbaths	96514	20770	22763	3143	753	1358	10882	5253	12883	932	2212	12076	1913

Appendix

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Input	Gover	Produc
Pretoria	7493481	34432	13759	1390701	113705	291617	690132	877260	1476068	172409	253480	2372334	314544
Soshanguve	15781	251	0	0	0	161	5090	109	5101	920	876	2568	2457
Wonderboom	465422	4778	1	374385	0	3432	16966	4872	7646	2529	1313	41154	10772
Johannesburg	12113772	13041	93731	2361481	259230	511656	3042487	1268299	3199946	450859	549515	1025858	436699
Randburg	1109056	10595	6971	304575	155633	106100	196648	54753	197834	23453	33973	63425	23042
Alberton	710059	940	0	401788	8065	29458	97383	28913	68695	8076	11797	64798	13740
Benoni	777111	5409	5516	313637	9060	41658	128043	44837	145343	11434	24959	81851	15282
Boksburg	1072913	-719	125640	545935	14222	88509	91448	54896	85248	6759	14639	60793	14821
Germiston	2535911	3093	4551	1240664	75168	210974	384289	236068	243661	22846	41843	129802	26633
Kempton Park	1859868	10425	2081	796220	9592	62782	316071	395203	198690	21250	32403	73061	16896
Brakpan	295704	1358	33292	102731	6743	14110	33102	13542	9861	4227	8563	37283	8018
Delmas	132099	62247	21619	12620	1047	2454	7738	4866	7825	1951	1344	9166	1910
Heidelberg	183130	14023	9967	102267	1633	5101	7037	6327	12184	2809	2092	22183	1691
Nigel	280408	13394	29059	118962	3832	9271	38304	6733	23133	2761	3972	35405	3526
Springs	1236120	227	105045	721300	10873	20386	109849	60464	113014	12610	19408	88159	13601

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Input	Gover	Produc
Krugersdorp	780131	25979	61223	336185	8595	18947	75209	40666	101592	9206	17446	107393	12582
Oberholzer	2370041	-9951	2211093	17529	2851	30203	42985	14179	35098	2252	6027	25327	4302
Randfontein	708699	3921	478214	94056	4176	16534	28674	18098	36218	2351	6220	26769	5908
Roodepoort	715350	-1366	96943	181684	14428	47744	87357	63087	131505	21983	22583	73384	21182
Westonaria	944335	3844	872055	8446	3362	8447	13865	5443	18765	719	3222	9263	3348
Brits	290071	60464	13076	115461	4271	6022	25467	5530	18179	2577	3122	37462	4684
Bronkhorstspruit	75563	26798	9649	4623	607	2646	7904	6542	4655	1639	799	5391	1708
Cullinan	72125	1373	56863	514	0	1144	3882	1197	2417	282	415	3222	1646
Vanderbijlpark	1094679	11862	179	786855	4960	40024	60588	23453	102828	7104	17658	52987	21477
Vereeniging	1139293	45581	9705	484187	122447	48403	135264	54733	135998	21372	23355	88347	14611
Sasolburg	783678	7743	58792	496360	87350	21987	24097	10842	53081	3254	9115	22677	6610
Gazankulu	207839	41101	3051	26395	0	18314	8881	8627	11589	539	1672	89905	1109
Kangwane	92215	11620	7234	7466	239	12312	1589	1546	11231	66	0	38445	467

Region	Total	Farm	Mine	Make	Supply	Build	Trade	Move	Bank	Serve	Imput	Gover	Produc
kwaNdebele	57845	7828	0	3836	0	8721	5923	769	4379	128	0	25781	430
kwaZulu	1084687	220930	467	174310	48	90157	35697	97448	97600	2864	1976	345626	21516
Lebowa	450593	24070	27225	49429	4403	39720	31270	24562	28630	1254	1095	202118	18907
Quaqua	92845	4068	0	11285	0	24508	2598	3290	16194	213	1850	32099	440

INDEX

(D) = Definition

25X1

Author: Kennedy Ian Geoffrey.

Name of thesis: Influence Of Gross Regional And Industrial Product Ranks On Data Call Connections.

PUBLISHER:
University of the Witwatersrand, Johannesburg