Serie Research Memoranda

Qualitative Data and Error Measurement in Input-Output Analysis

P. Nijkamp
J. Oosterhaven
H. Ouwersloot
P. Rietveld

Research Memorandum 1990-92
December 1990
Abstract

This paper is a contribution to the rapidly emerging field of qualitative data analysis in economics. Ordinal data techniques and error measurement in input-output analysis are here combined in order to test the reliability of a low level of measurement and precision of data by means of a stochastic method for transforming ordinal data into cardinal ones by using a minimum number of assumptions. The validity of the method is empirically tested by applying it to an existing regional input-output table for the Netherlands. It is concluded that the ordinal data method developed here gives a fairly reliable replication of the underlying quantitative input-output data.
1. Introduction

Qualitative data techniques and error measurement are somewhat related topics in economics. In applied economic research, analysts often face the problem of lack of reliable or precise data. Input-output analysis, which is very data demanding regarding inter-industry relationships, is a good example of this situation. Analysts usually employ quantitative point estimates of technical coefficients which suggest a higher degree of accuracy than is actually warranted. One may try to overcome this problem by providing an indication of the error possibly included in the data.

Another way of dealing with this measurement problem is to use qualitative (e.g., ordinal) data which allow experts to express their knowledge in a fairly flexible way. A major problem inherent in qualitative data is however, that they are difficult to handle for subsequent analytical purposes. Two approaches can in principle be distinguished. First, apart from just deleting qualitative data one may regard such data as quasi-quantitative data (although this is methodologically not justified); second, one may translate qualitative data into quantitative data by means of adjusted (often ad-hoc) approaches.

In this paper we present a method to link the concepts of qualitative data and error measurement by using a stochastic approach. More specifically, we develop a method for transforming in a consistent way ordinal data into cardinal ones - in the context of input-output analysis - by using a minimum number of assumptions on underlying probability distributions. At the heart of the stochastic method is a random number generator. By applying and repeating this method consecutively, an (empirical) distribution function of the data can be derived which generates amongst others standard errors, which can be used as a tool for error measurement.

The paper is organized as follows. Section 2 is devoted to a short review of qualitative data methods in economics. Section 3 then discusses error measurement in a part of economic research where non-cardinal data are often encountered, viz. input-output analysis. Section 4 describes the method in detail, while in section 5 it is applied to the construction of a regional input-output table for the Netherlands. Section 6 contains some retrospective remarks regarding this method, and provides some further reflections on its use in the field of I-O analysis.
2. Qualitative Data Analysis in Economics

The history of quantitative economic research is predominantly based on the presence of metric data, measured on a ratio or interval scale. In recent years, the insight has grown that several economic phenomena cannot always be meaningfully measured by means of a cardinal metric. This has led to an increased interest in the analysis of qualitative (e.g., ordinal) or categorical (e.g., dichotomous or polytomous) information. For example, in micro-economics and marketing science significant progress has been made in the treatment of disaggregate panel and longitudinal survey data (e.g., in discrete choice modelling; see for instance Manski and McFadden, 1981). But also in macro-economics various techniques (e.g., non-parametric statistics, multidimensional scaling analysis, sign-solvability analysis) have been developed in order to cope with the measurement of seemingly unmeasurable phenomena (see for a survey Nijkamp et al. 1985).

Clearly, the distinction between qualitative and quantitative information is not always unambiguous. There is rather a continuum of measurement scales, or as Lazarsfeld and Burton (1970, p.140) note: "there is a direct logical continuity from qualitative classification to the most rigorous forms of measurement, by way of intermediate devices of systematic ranking, ranking scales, multidimensional classifications, typologies and simple quantitative indices".

According to Roberts (1979), measurement theory should ideally provide a unique, intersubjective and consistent translation of observed objects into a set of logical symbols. Apart from the measurement errors (e.g., response errors in economic surveys), problems of measurement may emerge as a result of ambiguous definitions of variables, inappropriate data assembling methods, vague conceptualisations of phenomena (e.g., in terms of latent variables), or non-sound interpretations of results.

Recently developed methods for qualitative data analysis aim at taking into account the limitations inherent in measuring variables on a non-metric scale, while also avoiding non-permissible numerical operations on qualitative variables. In the meantime a wide variety of qualitative data techniques has been developed; in this section only a few examples will be mentioned.

It is interesting to observe that already back in 1947 Samuelson advocated the use of qualitative calculus, based on sign-solvability analysis, in order to analyse the structure of an economic model in
terms of the direction of influence (and its sign) caused by an exogenous impulse (see also Lancaster 1981 and Brouwer et al. 1989).

In the statistical analysis of association we have inter alia non-parametric measures of association (e.g., rank correlation measures), or multivariate data set methods (e.g., log-linear modeling of contingency tables, correspondence analysis, or multidimensional scaling).

In the field of statistical analysis of dependence there is also a diversity in methods, for example, generalised linear models, disaggregate choice methods, ordinal regression analysis, path analysis, qualitative LISREL methods, or partial least squares.

Major advances have also been made in explanatory behavioural modelling in economics. The rapid penetration of micro-based discrete choice models, based on categorical data, reflects a new interest in individual decision-making and utility evaluation. Examples of models in this framework are inter alia random utility models (e.g., logit, probit models, elimination-by-aspects models) or general extreme value models.

Special attention has to be given to the area of plan and project evaluation, where a wide variety of different multi-criteria evaluation methods has been developed for treating qualitative information, such as concordence methods, Saaty's prioritization method, and the regime analysis.

Finally, there are also significant advances regarding the mathematical treatment of non-metric data, for example, linguistic information in fuzzy set analysis, plausibility theory, or mixed qualitative calculus.

3. Errors in Input-Output Table Construction

Qualitative data are of essential importance because of the extreme cost of the construction of full survey input-output tables. Even at the national level, where far more data are readily available, full survey tables are rare. Most national tables either contain various non-survey elements (sometimes even entire rows and columns, especially for the service sector; see Uno 1989) or they represent updated tables (e.g., by inserting new or partial survey data in a RAS-context). Notwithstanding the obvious need for such information, it is very rare to find a quantified indication of the reliability of input-output tables (e.g., by means of standard errors). The same holds true for regional tables, where even less survey data are available and where most so-called
survey-based tables only have a partial or semi-survey character (see Stevens 1987).

Nevertheless, considerable effort has been made in investigating the sensitivity of input-output tables and the related multipliers for different types of errors. There are two types of such analyses. The first one uses deterministic approaches: the consequences of different non-survey assumptions are tested against the outcomes of a supposedly real survey and hence true table (see Richardson 1985, and Miller and Blair 1985 for recent overviews). The second one uses a probabilistic approach: density functions for input-output coefficients are either assumed or estimated, while next the deviations of the multipliers are analytically derived or simulated by means of Monte Carlo techniques (see Jackson and West 1989 for a recent overview).

The deterministic approaches focus usually on the question of the accuracy of non-survey techniques in time and space. In updating old (survey) tables, gradually consensus has been reached among input-output analysts about the use of RAS methods, provided that additional survey data about relatively large coefficients of the well-known A-matrix are available (see Lecomber 1975). Recently, the notion of large coefficients has been replaced by the more adequate concept of inverse-important input coefficients (see Sonis and Hewings 1989).

The construction of regional non-survey tables consists essentially of two steps. First, technical input coefficients are needed. Even if national technical coefficients ($a_{ij}^n$) are used to simulate the unknown regional ones ($a_{ij}^r$), it is preferable to use regional value added coefficients ($v_{ij}^r$, which are often known) and regional foreign import coefficients ($m_{ij}^r$, which are sometimes known) to rescale the national technical coefficients.

$$a_{ij}^r = a_{ij}^n \frac{(1 - v_{ij}^r - m_{ij}^r)}{(1 - v_{ij}^n - m_{ij}^n)}$$

where the upper index $r$ and $n$ refer to the region and the nation as a whole, respectively. Furthermore, (1) needs to be applied at the most disaggregated sectoral level in order to capture to a maximum extent the product-mix effect on the regional coefficients. Next, regional technical coefficients need to be corrected for domestic imports in order to derive the unknown regional input coefficients ($a_{ij}^{rr}$). When this correction is made in a multiplicative way, so-called regional purchase coefficients ($r_{ij}^r$) or RPC's are used, viz.:
In the case of regional input-output analysis most researchers agree on the non-acceptability of non-survey methods that are based on the minimization of interregional trade, whether explicitly (e.g., linear programming) or implicitly (e.g., location coefficients or supply-demand pool methods). Such methods are only applicable in case of homogeneous products traded between spatially separated point economies. They disregard essentially interregional cross-hauls, which are entirely rational in the presence of aggregated sectors, border trade and uniform delivered prices (cf. Richardson 1972). Thus they systematically overestimate the RPC’s and hence also all input-output multipliers.

Furthermore, it is interesting to note that Round (1978) using a biregional framework showed the inconsistency of the asymmetric use of location coefficients and supply-demand pool-methods in the case of positive net imports and positive net exports (see also Oosterhaven et al. 1986, for a general plea to use the biregional framework for the construction and updating of regional input-output tables).

The indirect estimation of the RPC’s via short-cut methods (cf. Katz and Burford 1981) or the RAS-method (Hewings 1977) does not provide a reliable solution either, as both approaches presuppose knowledge about the column sums (and in case of RAS also about the weighted row sums) of the $A^{TT}$-matrix, while this is precisely the type of information that is most difficult to obtain. Moreover, once this information is available it is more efficient to estimate the table in a more direct way.

Finally, in this case the interesting new concept of qualitative input-output analysis based on qualitative calculus (see Bon 1989) does not provide a solution either, mainly because in most developed regions the qualitative Leontief-inverse, which indicates whether or not a sector has backward linkages with another sector, will be entirely filled with plusses.\footnote{It is noteworthy that Bon (1989) draws also attention to the fact that the qualitative inverse for a supply-driven multiregional input-output model might be different from the corresponding demand-driven inverse. This outcome however, is only a conse-} Furthermore, no indication of the strength of the back-
ward linkages is given, whereas such information is crucial in all kinds of application.

Consequently, the only reasonable alternatives to full survey tables are hybrid tables (cf. Jensen and MacDonald 1982). Preferably such tables should use survey data on inverse important coefficients together with non-survey, econometric or other estimates of the RPC's (see also Stevens et al. 1989 for an econometric approach; and Leontief and Strout 1963 and Oosterhaven 1979 for the use of the gravity model).

Probabilistic analyses of errors in multipliers may provide information which is additional to the use of deterministic methods in analyzing the errors of non-survey methods. They do not replace such methods, as they presuppose or estimate density functions around the expected cardinal (non-survey) values. Their main purpose is to provide an indication of the confidence intervals of the deterministic estimates (cf. Lahiri and Satchell 1986; West 1986).

However, when one moves from the use of cardinal data to ordinal data, both alternative approaches may be integrated. An example of a deterministic approach is the method where one selects a pseudo-cardinal value for each coefficient, notwithstanding the fact that the analyst very often is only able to make a statement on whether, for instance, a certain RPC is significantly larger than another one. In such cases, often panels of experts have to be used in order to fix the values of such RPC's in the range of e.g., 0.00, 0.25, 0.50, 0.75 and 1.00 (see e.g. FNEI 1984a; 1984b; 1985). Clearly, this method does not imply that these values are the most reliable estimates of the true values concerned. It only ensures a reasonable fixation of coefficients according to five groups of values of increasing order of magnitude.

When the latter approach would be extended and satisfactorily formalized, one would shift from the use of cardinal data to ordinal data. Then one would be able to integrate the construction of a regional

(Footnote continued from previous page)

quence of the limited information on the destination of inter-regional trade in the multi-regional model. When full information is available, like e.g. in the standard Isard model (see Isard 1960) both qualitative inverses will again be equal and will most probably only be filled with plusses.
input-output table and the determination of the standard errors involved in this construction, which is essentially the aim of this paper.

4. Ordinal Data in Input-Output Analysis; a Stochastic Interpretation.

In this section we will discuss a method to deal with ordinal data in input-output analysis which naturally leads to an estimate of the error of the coefficients, thus combining the previously discussed themes.

4.1 Regional purchase coefficients

Suppose that ordinal information is available on the magnitude of RPC's, and that all sectors can be ranked in increasing order of the RPC's. We also suppose that an upper and lower limit (denoted as \( \bar{r} \) and \( \bar{r} \), respectively) are given for the RPC's. Then the set \( T \) of RPC's \( r_i \) which are consistent with the ordinal information can be presented as follows:

\[
T = \{(r_1, \ldots, r_I) | \bar{r} \leq r_1 \leq r_2 \leq \ldots \leq r_I \leq \bar{r}\}
\]  

(3)

The method adopted in this paper aims to use the information in (3) by means of a stochastic approach. This is done by introducing the probability that a certain combination of RPC's is the true combination consistent with the ordinal information. In this case the 'principle of insufficient reason', which forms also the basis for the well-known Laplace criterion in case of decision-making under uncertainty is used (see Taha, 1976). This principle states that if qualitative information is emerging from an (unknown) quantitative data set, there is a priori and without additional information no reason to assume that any value has a higher probability. Therefore a rectangular (or uniform) distribution is the most plausible one. Assuming in our case no further additional information, the uniform probability distribution is thus the most logical one to use, as it assumes that all elements in \( T \) are equally probable. This gives rise to the following distribution:


\[ g(r_1, \ldots, r_I) = \frac{1}{1/(I-1) - r} \quad \text{if} \quad r_1 \leq r_1 \leq r \]
\[
\begin{align*}
& r_1 \leq r_2 \leq r \\
& \quad \vdots \\
& r_{I-1} \leq r_I \leq r
\end{align*}
\]

\[ = 0 \quad \text{elsewhere} \]

In our analysis we will then use a random generator for drawing numerical values from this distribution. Appendix I contains a description of the procedure for generating random combinations of coefficients which are consistent with (4). An analytical formulation of expected values of such coefficients is given in Rietveld (1984).

4.2 Input coefficients

For input coefficients a similar approach can be followed although there is one difference: the coefficients \( a_i \) (including both intermediate and primary inputs) in a certain column of the input-output matrix should satisfy the constraint that they add up to 1. Thus, if we assume that the sectors have been ranked in increasing order of input coefficients \( a_i \), the set \( S \) of all coefficients satisfying the ordinal information is:

\[ S = \left\{(a_1, \ldots, a_I) \mid 0 \leq a_1 \leq a_2 \leq \ldots \leq a_I; \sum a_i = 1\right\}, \quad (5) \]

where the index of the column number of the input coefficients has been dropped for notational convenience. If we assume again that all points in \( S \) are equally probable, we find the following density function:

\[ f(a_1, \ldots, a_I) = c \quad \text{if:} \quad 0 \leq a_1 \leq 1/I \]
\[
\begin{align*}
& a_1 \leq a_2 \leq 1/(I-1) - a_1/(I-1) \\
& \quad \vdots \\
& a_{I-2} \leq a_{I-1} \leq 1/2 - a_{I-2}/2 - \ldots - a_{I-2}/2
\end{align*}
\]

\[ = 0 \quad \text{elsewhere} \]
where \( c \) can be shown to be equal to \((I-1)!1!\) (Rietveld, 1989). Once the values of \( a_1, \ldots, a_{I-1} \) are known, the value of \( a_I \) can be found as:
\[
1 - a_1 - \cdots - a_{I-1}.
\]

An operational approach to generate a random sample of coefficients consistent with (6) is also given in Appendix I. Analytical expression for the expected values of the input coefficients are given in Rietveld (1989).

### 4.3 Ranking with degrees of difference

Consider ordinal information such as \( x_1 < x_2 \) and \( x_2 < x_3 \). So far we have assumed that the degree of difference between \( x_1 \) and \( x_2 \) is equal to that between \( x_2 \) and \( x_3 \). In certain cases, analysts may be able to express their opinions in terms of rankings with varying degrees of difference. For example, \( x_1 \) is smaller than \( x_2 \) which is in turn considerably smaller than \( x_3 \). Information of this type is used in the analytical hierarchy process developed by Saaty (1977).

We will show that it is possible to develop the stochastic approach of section 4.2 in such a way that it can also deal with rankings of input coefficients or RPC's with varying degrees of difference. Assume that the degree of difference can be indicated by an index number \( m = 1, 2, \ldots \), where we assume: the higher \( m \) the larger the difference. Then the following notation will be used:

\[
x \leq_m y \quad \text{x is smaller than y according to degree } m
\]

where \( m = 1, 2, 3, \ldots \). In our stochastic approach variations in the degree of difference are taken into account by introducing auxiliary variables. For example, when \( x \leq_2 y \), an auxiliary variable \( b \) is added such that \( x < b < y \). Similarly when \( x \leq_3 y \), two auxiliary variables \( b \) and \( c \) are added such that \( x < b < c < y \).

Consider now the case of ranked information on RPC's. The following notation will be used:

\[
\tilde{r} \leq_{m_1} r_1 \leq_{m_2} r_2 \cdots \leq_{m_I} r_{I-1} \leq_{m_{I+1}} \tilde{r} \quad (7)
\]

where \( m_i \) is the degree of difference between \( r_{i-1} \) and \( r_i \) for \( i=1, \ldots, I \). Furthermore, let \( q_{ik} \) denote the \( k \)'th auxiliary variable between \( r_{i-1} \) and
For notational convenience we set $q_{im} = r_i$. Then (7) is equivalent to:

$$\tilde{r} \leq q_{11} \leq \ldots \leq q_{1m_1} \leq q_{21} \leq \ldots \leq q_{2m_1} \leq \ldots \leq q_{I+1,1} \leq \ldots \leq q_{I+1,m_{I+1}-1} \leq \tilde{r} \tag{8}$$

For our stochastic approach this means that instead of the original $I+1$ variables now $L = (\sum m_i) - 1$ variables have to be generated. The modified probability density function then reads as:

$$g(q_{11}, \ldots, q_{I+1, m_{I+1}-1}) = \frac{L!}{(\tilde{r} - \tilde{r})^L} \tag{9}$$

for all $q_{ik}$ satisfying (8).

It should be noted that from the generated values only the $q_{im_i} (= r_i)$ are used in the subsequent analysis. Clearly, the structures of (9) and (4) are rather similar; if $m_i = 1$ for all $i$, (9) and (4) coincide. A high degree of difference between subsequent RPC's can be shown to lead to a small variance of the $r_i$. For expected values of the coefficients a different result can be proven: if the degree of difference is the same for all subsequent RPC's ($\mu_1 = \mu_2 = \ldots = \mu_I = N$), the expected values of the coefficients do not depend on $N$.

Next, for input coefficients, the introduction of varying degrees of differences leads to a more complex adjustment of the original formulations. Using the same notation as above, our point of departure is:

$$0 \leq a_1 \leq a_2 \leq \ldots \leq a_I$$

$$\sum a_i = 1 \tag{10}$$

We then introduce auxiliary variables $\mu_{ik} (k-1, \ldots, m_i-1)$ satisfying the following condition (writing $a_i = \mu_{im_i}$ for notational convenience):
It can be shown that the modified probability density function \( f \) is equal to:

\[
f(\mu_{11}, \ldots, \mu_{I}, m_{I-1}) = \left( \sum_{i=1}^{I} m_i \right)! \prod_{i=1}^{I-1} \left( \frac{m_i}{i+1} \right)\tag{12}
\]

for all \( \mu_{ik} \) satisfying (11). Note that when \( m_i = 1 \) for all input coefficients, (12) coincides with (6). The approach described in Appendix I can after some modifications be used again to generate random values consistent with (12).

### 4.4 Ties and incomplete rankings

Ties deserve special attention in ordinal data; the probability that ties occur is high in the case of a large number of ranked observations. Consider, for instance, the following ranking: \( a_1 \leq (a_2, a_3) \leq a_4 \).

This may have different interpretations:

- \( a_2 \) and \( a_3 \) are exactly equal
- \( a_2 \) and \( a_3 \) are approximately equal
- \( a_2 \) and \( a_3 \) are incomparable: \( a_2 \) may be either larger or smaller than \( a_3 \), and the difference between the two is not necessarily small.

Each of these cases deserves its own treatment in the stochastic approach outlined above.

When observations are exactly equal, one only needs to draw one random value which is assigned to all observations concerned. An inspection of (9) and (12) reveals that this can be done in a consistent way by interpreting an exact equality as \( \leq_0 \) (i.e., \( m_i = 0 \) in such a case). Thus, there is no need to design special procedures to deal with exact equality: one can still use the formulas derived in Appendix I.

In the case of incomparable observations (an incomplete ranking) one can still use the stochastic approach. Consider for example a cluster of incomparable observations consisting of \( a_2, a_3 \) and \( a_4 \). Then random numbers \( x \leq y \leq z \) are generated which are assigned to \( a_2, a_3 \) and \( a_4 \).
in a random way. Thus, in one case \(a_2\) may be assigned the largest value \(z\), and in another case the smallest value.

In the case of approximately equal observations, one may proceed as follows. In a first step, a value is generated for these observations as if they were exactly equal (along the lines sketched above). In the second step, the observations are assumed to be uniformly distributed in an appropriately defined interval around this value. Consider for example three clusters: \(\{r_1\}, \{r_2, r_3\}\) and \(\{r_4, r_5\}\) with the following features: \(r_1 \leq r_2 = r_3 \leq r_4 = r_5\), where \(r_2 = r_3\) means that \(r_2\) and \(r_3\) are approximately equal. The standard stochastic approach leads to values \(b_1, b_2\) and \(b_3\) for the three clusters. Then in the last step, values for \(r_2\) and \(r_3\) are drawn from a uniform distribution on the interval \([\frac{1}{2} b_1 + \frac{1}{2} b_2, \frac{1}{2} b_2 + \frac{1}{2} b_3]\). This approach can also be followed for input coefficients, but in that case an additional condition \((r_2 + r_3 = 2b_2)\) has to be imposed to ensure that the additivity constraint on the coefficients is satisfied.

Thus in all cases the probability approach is in principle applicable and hence we may conclude that the stochastic approach outlined above is quite flexible. It can deal with all kinds of ordinal data on coefficients with and without additivity constraints:
- standard ordinal data
- ordinal data with degrees of difference
- ties of exactly equal observations
- ties of approximately equal observations
- incomparable observations.

Now the empirical question has to be answered whether the above mentioned approach leads to reliable results. This will be done in the next section.

5. Empirical Results

The best way of testing empirically the reliability and appropriateness of the ordinal input-output data method is to use an existing quantitative input-output table, to transform it into ordinal rankings of the coefficients and next to investigate whether our ordinal data method is able to regenerate with a high degree of reliability the information contained in the original table.

In our case the appropriateness of the method will be tested by means of an existing regional input-output table. The table concerned is that of the province of Groningen in the Northern part of the
Netherlands for the year 1980. The table is given in FNEI (1986) and contains 9 sectors, one of which is the public sector. Household consumption is also given which - in combination with the availability of figures on wages and salaries as a fraction of value added - offers the possibility of endogenizing the household sector.

Although this input-output table is published in quantitative terms, a main part of it has a soft empirical basis. We have used types of ordinal data as described in section 4 to represent the essential characteristics of the underlying information. In the second step we have used the stochastic approach of section 4 to generate random sets of quantitative values consistent with these ordinal data. The different versions of the input-output tables are next compared by means of the indirect income multipliers they generate. As we have seen in section 4 ordinalization can be carried out in various ways. Therefore two related questions may be asked: how well are multipliers estimated when all ordinal information is incorporated in the procedure and how do the various ways of formulating ordinal data affect the numerical results? Both questions will be dealt with, but first we will discuss the ordinalization phase itself.

The ordinalization we used is based on national input-output coefficients together with regional purchase coefficients (RPC's). This reflects the usual way regional tables are constructed. We assume here that quantitative values of the value added coefficients are known for each sector. Also some cells that by definition are zero in the original table are set equal to zero in the estimated table at the outset. As concerns the RPC's, we make a distinction between two vectors of RPC's, one of which applies to the input coefficients and another one to household consumption. We assume the availability of information on the upper and lower bounds on both RPC vectors. RPC's that are known to be 1 are also assigned that value exogenously. Finally, wages and salaries as a fraction of value added are also assumed to be known in a quantitative way. Thus, the stochastic approach is applied to the input-output coefficients, including foreign imports, consumption coefficients and the

1 We have decided to evaluate the differences only by means of the indirect part of the multiplier, as the direct part is exogenously known. Comparing total multipliers underestimates the size of the differences at hand (cf. Oosterhaven et al, 1986)
RPC vectors. The results of our stochastic approach are summarized by the mean values of the multipliers.

When we want to compare the indirect income multipliers based on the ordinal data with the ones implied by the initial cardinal data, two possibilities can be distinguished, depending on the way in which the RPC's for the input coefficients are used. Starting with the national input-output table, the first possibility is multiplying each single element with a separate RPC which leads to the given regional input-output table. The second possibility is multiplying each element of a row with a common RPC. This means that in equation (2) \( r_{ij}^F = r_{ij}^C \) for all \( j \). The first approach is obviously rather information demanding. The second approach is more common practice but leads to less accurate multipliers.

In our processing of the ordinal information we have only considered the case of one RPC per sector (row), as the treatment of RPC's per cell is demanding too much information to be of any practical value in the ordinal data approach. Of course, it is interesting to know what the influence is of this simplification on our results, i.e., the sensitivity of using sectoral RPC's per cell. Then it is also important to know whether biases resulting from our ordinal data method show similar patterns to the biases caused by the use of a sectoral RPC.

So we will first investigate the bias caused by the use of a sectoral RPC by comparing multipliers based on the given input-output table; see Table 1. This table shows multipliers from the original regional input-output table, in which household consumption is treated both exogenously and endogenously (columns 1 and 4, respectively). Columns 2 and 5 show then the respective multipliers which result when the national input coefficients are multiplied by sectoral RPC's. Next columns 3 and 6 show the bias in columns 2 and 5 as a percentage of column 1 and 4, respectively.
Table 1  Indirect income multipliers with RPC per sector and per cell

<table>
<thead>
<tr>
<th>Sector</th>
<th>Household consumption</th>
<th>Household consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>exogenous</td>
<td>endogenous</td>
</tr>
<tr>
<td></td>
<td>Cell</td>
<td>Sectoral</td>
</tr>
<tr>
<td></td>
<td>RPC</td>
<td>RPC</td>
</tr>
<tr>
<td>Agriculture</td>
<td>.112</td>
<td>.092</td>
</tr>
<tr>
<td>Industry</td>
<td>.468</td>
<td>.491</td>
</tr>
<tr>
<td>Utilities</td>
<td>.651</td>
<td>.462</td>
</tr>
<tr>
<td>Construction</td>
<td>.377</td>
<td>.353</td>
</tr>
<tr>
<td>Trade</td>
<td>.042</td>
<td>.045</td>
</tr>
<tr>
<td>Transport</td>
<td>.077</td>
<td>.077</td>
</tr>
<tr>
<td>Comm. services</td>
<td>.153</td>
<td>.150</td>
</tr>
<tr>
<td>Other services</td>
<td>.072</td>
<td>.066</td>
</tr>
<tr>
<td>Public sector</td>
<td>.094</td>
<td>.094</td>
</tr>
</tbody>
</table>

In general the bias resulting from the use of a sectoral RPC is small, except for the sectors 1 and 3 (in case of exogenous household consumption) and sector 3 (in case of endogenous household consumption). More generally one observes a reduction in errors, when household consumption, being less sector-specific than intermediate demand, is treated as an endogenous variable.

Table 2 shows the outcomes of the ordinal estimation procedure when all types of ordinal information are used and when household consumption is kept exogenous. The multipliers in the first column are those resulting from the original cardinal input-output table (cf. Table 1). The second column shows the mean value of the multipliers resulting from using the ordinal data method 500 times, with their standard deviation in parentheses. Column 3 shows the bias of (2) as a percentage of (1).
Table 2. Indirect income multipliers with exogenous household consumption.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Initial (n = 500)</th>
<th>Ordinal data method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Relative</td>
<td>mean stand</td>
<td>difference</td>
</tr>
<tr>
<td></td>
<td>dev.</td>
<td></td>
</tr>
<tr>
<td>1. Agriculture</td>
<td>.112</td>
<td>.099</td>
</tr>
<tr>
<td>2. Industry</td>
<td>.468</td>
<td>.499</td>
</tr>
<tr>
<td>3. Utilities</td>
<td>.651</td>
<td>.672</td>
</tr>
<tr>
<td>4. Construction</td>
<td>.377</td>
<td>.415</td>
</tr>
<tr>
<td>5. Trade</td>
<td>.042</td>
<td>.057</td>
</tr>
<tr>
<td>6. Transport</td>
<td>.077</td>
<td>.083</td>
</tr>
<tr>
<td>7. Comm. services</td>
<td>.153</td>
<td>.155</td>
</tr>
<tr>
<td>8. Other services</td>
<td>.072</td>
<td>.068</td>
</tr>
<tr>
<td>9. Public sector</td>
<td>.094</td>
<td>.090</td>
</tr>
</tbody>
</table>

Table 2 shows that the standard deviations as generated by the ordinal approach are rather large, in most cases (far) over 10% of the respective means. All initial multipliers are within the interval of mean + or - standard deviation. More interesting is the fact that the ranking of the multipliers with respect to their magnitude is the same in both cases. This means, for instance, that the sector with the largest multiplier effect is correctly identified. The biases do not follow a clear pattern, although we note that really large biases (over 20%, for instance) are rare. When we compare column (4) of Table 2 with column (3) of Table 1, there is no indication that the biases in our qualitative data method are a consequence of the use of sectoral RPC.

Table 3 shows multipliers when household consumption is made endogenous. Again all types of ordinal information are used.
Table 3. Indirect income multipliers with endogenous household consumption.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Initial mean (1)</th>
<th>Ordinal data method mean (2)</th>
<th>Ordinal data method stand dev. (3)</th>
<th>Relative difference between (1) and (2) in % (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Agriculture</td>
<td>0.452</td>
<td>0.456</td>
<td>0.054</td>
<td>0.9</td>
</tr>
<tr>
<td>2. Industry</td>
<td>0.917</td>
<td>1.003</td>
<td>0.181</td>
<td>9.4</td>
</tr>
<tr>
<td>3. Utilities</td>
<td>1.157</td>
<td>1.229</td>
<td>0.198</td>
<td>6.2</td>
</tr>
<tr>
<td>4. Construction</td>
<td>0.799</td>
<td>0.864</td>
<td>0.091</td>
<td>8.1</td>
</tr>
<tr>
<td>5. Trade</td>
<td>0.362</td>
<td>0.399</td>
<td>0.050</td>
<td>10.2</td>
</tr>
<tr>
<td>6. Transport</td>
<td>0.407</td>
<td>0.438</td>
<td>0.054</td>
<td>7.6</td>
</tr>
<tr>
<td>7. Comm. services</td>
<td>0.507</td>
<td>0.529</td>
<td>0.051</td>
<td>4.3</td>
</tr>
<tr>
<td>8. Other services</td>
<td>0.401</td>
<td>0.415</td>
<td>0.046</td>
<td>3.5</td>
</tr>
<tr>
<td>9. Public sector</td>
<td>0.429</td>
<td>0.445</td>
<td>0.047</td>
<td>3.7</td>
</tr>
</tbody>
</table>

The results of Table 3 are quite similar to those of Table 2. Again almost all standard deviations appear to be over 10% of the means and again all initial multipliers are within a range of mean ± standard deviation. Column (4) of Table 3 and column 6 of Table 1 do not display a similar pattern. The only remarkable result of Table 3 is that all initial multipliers are (slightly) overestimated by the ones of the ordinal data method. There is no intrinsic reason why this should be the case. It is an accidental result of the way the initial coefficients have been reformulated in ordinal terms. Finally, we note that - even though there is a cluster of initial multipliers which have a similar order of magnitude - the ranking of the multipliers is again correctly 'predicted' by the ordinal data method.

Next we discuss the various stages of adding new information on our data. In Table 4 the following cases are distinguished. In the base case the ordinalization is carried out without allowing ties or degrees of differences. In the "=" case the base case is extended with the allowance of ties which are all interpreted as 'equal to'. The "=" case is similar to the "=" case, but now with ties interpreted as 'approximately
equal to'. In the 'both' case both interpretations of ties are allowed for. In the case labelled 'DOD' again no ties are allowed, but in that case we make use of the concept of degree of difference. Finally, in the case of label 'ALL', all types of ordinal information are used simultaneously.

In the present context where quantitative guesses of coefficients are already available, the interpretation of ties as being incomparable is not relevant, and hence we decided to drop this interpretation in the experiments reported here.

Table 4. Effect of extra information on mean values of multipliers (initial multiplier: = 100); household consumption endogenous; (n=500).

<table>
<thead>
<tr>
<th>Sector</th>
<th>Base</th>
<th>=</th>
<th>\approx</th>
<th>Both</th>
<th>DOD</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Agriculture</td>
<td>108</td>
<td>105</td>
<td>104</td>
<td>104</td>
<td>105</td>
<td>101</td>
</tr>
<tr>
<td>2. Industry</td>
<td>136</td>
<td>127</td>
<td>127</td>
<td>128</td>
<td>119</td>
<td>109</td>
</tr>
<tr>
<td>3. Utilities</td>
<td>140</td>
<td>135</td>
<td>134</td>
<td>136</td>
<td>109</td>
<td>106</td>
</tr>
<tr>
<td>4. Construction</td>
<td>112</td>
<td>111</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>108</td>
</tr>
<tr>
<td>5. Trade</td>
<td>122</td>
<td>117</td>
<td>117</td>
<td>117</td>
<td>118</td>
<td>110</td>
</tr>
<tr>
<td>6. Transport</td>
<td>108</td>
<td>104</td>
<td>104</td>
<td>103</td>
<td>112</td>
<td>108</td>
</tr>
<tr>
<td>7. Comm. services</td>
<td>100</td>
<td>97</td>
<td>97</td>
<td>96</td>
<td>107</td>
<td>104</td>
</tr>
<tr>
<td>8. Other services</td>
<td>98</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>106</td>
<td>103</td>
</tr>
<tr>
<td>9. Public sector</td>
<td>100</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>107</td>
<td>104</td>
</tr>
</tbody>
</table>

mean absolute percentage error

14 12 12 12 10 6

Table 4 shows some interesting results. First, the way in which ties are interpreted does not make any significant difference for the results. Second, there is a significant difference in the dispersion in index values of the DOD and ALL cases compared to the other cases. Third, the ALL case appears to give the best estimates, although only with respect to the average of the index numbers and not for the last three sectors. We will discuss these results in more detail below.

From the discussion in section 4 it can be derived that the way in which a tie is interpreted does not make any difference for the mean value of a single element which is drawn. At the same time it appears to be impossible to derive analytically the distribution function of the
multipliers involved because of the stochastically complex operations of multiplication and inversion that have to be carried out in order to arrive at the multipliers. Table 4 now shows that the mean values may be considered identical (the small differences can be attributed to sampling effects), while Table 5 shows that also the standard deviations are very much similar (for the sake of completeness the standard deviations of the other cases are also given).

Table 5. Effect of extra information on standard deviations of multipliers; household consumption endogenous; (n=500).

<table>
<thead>
<tr>
<th>Sector</th>
<th>base</th>
<th>=</th>
<th>=</th>
<th>both</th>
<th>DOD</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Agriculture</td>
<td>.065</td>
<td>.070</td>
<td>.066</td>
<td>.068</td>
<td>.056</td>
<td>.054</td>
</tr>
<tr>
<td>2. Industry</td>
<td>.183</td>
<td>.227</td>
<td>.224</td>
<td>.212</td>
<td>.170</td>
<td>.181</td>
</tr>
<tr>
<td>3. Utilities</td>
<td>.287</td>
<td>.311</td>
<td>.308</td>
<td>.296</td>
<td>.196</td>
<td>.198</td>
</tr>
<tr>
<td>4. Construction</td>
<td>.114</td>
<td>.114</td>
<td>.119</td>
<td>.117</td>
<td>.091</td>
<td>.091</td>
</tr>
<tr>
<td>5. Trade</td>
<td>.061</td>
<td>.067</td>
<td>.064</td>
<td>.067</td>
<td>.051</td>
<td>.050</td>
</tr>
<tr>
<td>6. Transport</td>
<td>.065</td>
<td>.065</td>
<td>.063</td>
<td>.066</td>
<td>.054</td>
<td>.054</td>
</tr>
<tr>
<td>7. Comm. services</td>
<td>.062</td>
<td>.064</td>
<td>.063</td>
<td>.064</td>
<td>.052</td>
<td>.051</td>
</tr>
<tr>
<td>8. Other services</td>
<td>.057</td>
<td>.058</td>
<td>.056</td>
<td>.058</td>
<td>.048</td>
<td>.046</td>
</tr>
<tr>
<td>9. Public sector</td>
<td>.058</td>
<td>.060</td>
<td>.058</td>
<td>.060</td>
<td>.048</td>
<td>.047</td>
</tr>
</tbody>
</table>

Thus in an empirical sense the distinction between the various ways of dealing with ties is apparently not necessary. This means that all ties may be considered as 'equal to' ties which are computationally easier to handle.

We checked the above statement by generating multipliers allowing for degrees of differences, by combining them with various interpretations of ties. The hypothesis that the various ways of dealing with ties are redundant leads to the expectation that the cases of DOD plus "=" and DOD plus "≡" both give multipliers more or less identical to the 'ALL' case. It turned out that indeed the mean values of the multipliers were in all three cases almost identical (the difference in index numbers was 1 point at most), while also the standard deviations were very much alike.

Returning to the results of Table 4, we already noted the difference in dispersion in the index numbers of the 'ALL' case compared to the other cases. Of course a situation where all multipliers are
estimated with the same error is preferable to a case where some multipliers are estimated without any error and some others with rather large errors, especially when it is not known which multipliers are wrongly estimated. This observation makes clear that all available information should be incorporated when using the ordinal data method. But examining the underlying multipliers reveals an even more important result, viz. that the ranking of the multipliers is only correctly generated by the 'ALL' case (cf. Table 3). This statement was checked by also examining in an analogous way the multipliers resulting from the various cases of the information used, when household consumption was kept exogenously. Also these results appear to confirm the above conclusion. Only in the case when all information is used, the ranking of the multipliers generated is the same as in the initial case. Thus we may conclude that the ALL case is the only one giving a correct picture of the quantitative structure of the economy.

6. Conclusion

Quantitative economic analysis aims at deriving conclusions with a maximum degree of reliability. If in a given case no cardinal data are available nor can be obtained in the framework of a certain research project and if instead ordinal data are available, then good practice means to derive as much information as possible from the existing data set. The ordinal data method described in this paper and applied in the framework of input-output analysis has clearly shown that this is a meaningful endeavour which - by generating cardinal figures out of ordinal data - generates numerically plausible results. In the test case, where original and re-generated cardinal data are compared, it appears that the point estimates are satisfactory when all available information is used. The errors of an ordinal input-output analysis appear to be relatively small, for instance, in comparison to errors when alternative updating procedures are used (see Oosterhaven 1986).

In input-output analysis one of the ways often used to overcome the problem of missing quantitative data is to use expert judgements to arrive at cardinal values of the coefficients concerned. The reliability of the ensuing multipliers is questionable however. The ordinal approach presented here is an attractive alternative since the standard deviations it generates are a useful means to judge the range of uncertainty on the multipliers because of data weaknesses. The ordinally estimated multipliers however, should not (or not necessarily) replace
the cardinal ones. Also, it is preferable to use hard (cardinal) information whenever possible, very much in the same way as we used the value added coefficients. Our method is able to cope with this kind of mixed data.

Concerning the technical part of the method, the interesting result emerges that in this case study the interpretation of ties ('equal to' versus 'approximately equal to') is not important. On the other hand, the concept of degree of differences appears to be essential for the use of our method.

The main advantage of ordinal input-output analysis is the integration of input-output table construction and the determination of standard errors into one framework. Our analysis has demonstrated - based on an empirical test case - that ordinal input-output analysis may offer an extremely valuable alternative in case of missing or imprecise information on technical coefficients in an input-output model.
1. Generating regional purchase coefficients

As indicated by Mood and Graybill (1963), it is not difficult to generate random values for $r_1, \ldots, r_I$. They show by means of order statistics that one can start with drawing $I$ numbers from the uniform distribution on $[0,1]$, after which $r_1$ is assigned the smallest number, $r_2$ the one but the smallest number, etc.

An alternative approach would be the following one. Taking (4) as a starting point, it can be shown that the marginal distribution of the smallest RPC reads as follows:

$$g(r_1) = \frac{1}{I} \left( \frac{\bar{r} - r_1}{\bar{r} - \hat{r}} \right) I \hat{r} \leq r_1 \leq \bar{r}$$

$$= 0 \text{ elsewhere}$$

Further, the conditional density functions can be shown to read as follows for $i = 2, \ldots, I$:

$$g(r_i|r_1, \ldots, r_{i-1}) = \frac{(I-i+1)(\bar{r} - r_i)}{(I-r_{i-1})} I-i+1 \leq r_i \leq \bar{r}$$

where $r_{i-1} \leq r_i \leq \bar{r}$

Thus, a random vector with RPC's can be generated by drawing a value for $r_1$ on the basis of $g(r_1)$, followed by drawing a value for $r_2$ on the basis of $g(r_2|r_1)$, etc. However, these conditional distributions are not included in standard statistical packages. Therefore, random values cannot be directly created by means of random generators. A solution for this problem is given by the theorem which says that if $G(x)$ is the distribution of $x$, then $u = G(x)$ is uniformly distributed on the interval $0 \leq u \leq 1$ (Hogg and Craig, 1970, p. 349). For the latter uniform distribution, standard random generators are available. Then if $u_1$ is uniformly distributed on the interval $[0,1]$, $r_1 = G^{-1}(u_1)$ can be shown to be distributed according to the density function $g(r_1)$. Thus, random values for $r_1$ can be found by using the following transformation:
For \( i = 2, \ldots, I \), the following transformation has to be used:

\[
    r_i = r - (r - r_{i-1}) (1 - u)^{1/(I-i+1)}
\]

2. Generating input coefficients

The starting point is the joint density function of input coefficients:

\[
    f(a_1, \ldots, a_{I-1}) = c \begin{cases} 
        0 \leq a_1 \leq 1/I \\ 
        a_1 \leq a_2 \leq (1-a_1)/(I-1) \\ 
        a_2 \leq a_3 \leq (1-a_1-a_2)/(I-2) \\ 
        \vdots \\ 
        a_{I-2} \leq a_{I-1} \leq (1-a_1, \ldots, a_{I-2})/2 \\ 
    \end{cases} 
\]

= 0 \text{ elsewhere}

where \( c = (I-1)!I! \)

Then the marginal density function of \( a_1 \) can be derived as:

\[
    f(a_1) = (I-1)I(1-Ia_1)^{I-2} \text{ for } 0 \leq a_1 \leq (1/I) \\
    = 0 \text{ elsewhere}
\]

Furthermore, the conditional density functions can be shown to read as follows for \( i = 2, \ldots, I-1 \)

\[
    f(a_i | a_1, \ldots, a_{i-1}) \\
    = (I-i)(I-i+1) \frac{[1-a_1-\cdots-a_{i-1}-(I-i+1)a_i]^{I-i-1}}{[1-a_1-\cdots-(I-i+2)a_{i-1}]^{I-i}}
\]

where \( a_{i-1} \leq a_i \leq (1-a_1-\cdots-a_{i-1})/(I-i+1) \)

Then a random vector with input coefficients can be generated by drawing a value for \( a_1 \) on the basis of \( f(a_1) \), followed by drawing a value for \( a_2 \).
on the basis of \( f(a_2|x_1) \), etc. Finally, \( a_1 \) can be computed as \( 1-a_1 \) \( \ldots \) \( a_{I-1} \).

Let again \( u_i \) be a number drawn from the uniform distribution on the interval \([0,1]\). Then it can be shown that the following transformation has to be used to generate random values of \( a_1 \):

\[
a_i = \frac{1}{i} \left[ 1 - \left( 1 - u_i \right)^{1/(I-1)} \right]
\]

For \( a_2, \ldots, a_{I-1} \) the following transformation has to be used:

\[
a_i = \left[ (1-a_{i-1}) (1-a_1) \ldots (1-a_{i-2}) (1-u_i)^{1/(I-1)} \right] / (I-1)
\]

Finally, \( a_I \) can be computed as \( 1-a_1 \) \( \ldots \) \( a_{I-1} \).
References


| 1989-2 | J.C. van Ours | Self-Service Activities and Legal or Illegal Market Services |
| 1989-3 | H. Visser | The Monetary Order |
| 1989-5 | N.M. van Dijk | A Simple Throughput Bound for Large Closed Queueing Networks with Finite Capacities |
| 1989-6 | N.M. van Dijk | Analytic Error Bounds for Approximations of Queueing Networks with an Application to Alternate Routing |
| 1989-7 | F. Spreij | Self-exciting Counting Process Systems with Finite State Space |
| 1989-8 | H. Visser | Rational Expectations and New Classical Macroeconomics |
| 1989-9 | J.C. van Ours | De Nederlandse boekenmarkt tussen stabiliteit en verandering |
| 1989-10 | H. Tieleman A. Leliveld | Traditional "Social Security Systems" and Socio-economic Processes of Change: The Case of Swaziland; opportunities for research |
| 1989-11 | N.M. van Dijk | "Stop - Recirculate" for Exponential Product Form Queueing Networks with Departure Blocking |
| 1989-12 | F.A.G. den Butter | Modelbouw en metingsbeleid in Nederland |
| 1989-13 | N.M. van Dijk | Simple performance estimates and error bounds for slotted ALOHA loss systems |
| 1989-14 | H. Clemens J.P. de Groot | Sugar Crisis, a Comparison of two Small Peripheral Economies |
| 1989-15 | I.J. Steyn | Consistent Diffuse Initial Conditions in the Kalman Filter |
| 1989-16 | I.J. Steyn | Ase Estimation of Parameters in a State Space Model |
| 1989-17 | B. Vogelvang | Dynamic Interrelationships between Spot Prices of some Agricultural Commodities on Related Markets |
| 1989-18 | J.C. van Ours | Zoeken naar nieuwe medewerkers |
| 1989-19 | H. Kox | Integration of Environmental Externalities in International Commodity Agreements |
| 1989-20 | P.H.F.H. van Casteren A.H.Q.H. Nerbies | Micro Labour Demand Functions with Heterogeneous Output for Dutch Housing-Construction |
| 1989-21 | J.C. van Ours | An empirical Analysis of Employers' Search |
| 1989-22 | R.J. Boucherie H.H. van Dijk | Product Form for Queueing Networks with State Dependent Multiple Job Transitions |
| 1989-23 | N.M. van Dijk | On "stop - repeat" Servicing for Non-exponential Queueing Networks with Blocking |
| 1989-25 | A.F. de Vos | Kansen en risicomen - Over de fundamenten van statistische uitspraken door accountants |
| 1989-26 | N.M. van Dijk | A Note on Extended Uniformization for Non-Exponential Stochastic Networks |
| 1989-28 | N.M. van Dijk F.J.J. Trapman | Exact Solutions for Central Service Systems with Breakdowns |
| 1989-29 | H.M. van Dijk | Product Form for Queueing with Limited Clusters |
| 1989-30 | A. Farrel | Tijsinngeld van huishoudelijke activiteiten in relatie tot kenmerken van huishoudens |
| 1989-32 | N.M. van Dijk | A Simple Performatibility Estimate for Jackson Networks with an Unreliable Output Channel |
| 1989-34 | N.M. van Dijk | An Equivalence of Communication Protocols for Interconnection Networks |
| 1989-35 | H. Visser | Micro-Foundations of Money and Finance |
| 1989-36 | N.M. van Dijk | The Importance of Bias-Terms for Error Bounds and Comparison Results |
| 1989-37 | A.F. de Vos | On Regression Sampling in Statistical Auditing: And Answers to the Wrong Questions? |
| 1989-38 | R.J. Huiskamp | Company Strategy and the (Re)Design of Industrial Relations, some case studies in the Netherlands |
| 1989-39 | Dr. Raul Ruben | Notes sobre la cuestion agraria en Honduras |