Environment and Planning A 2003, volume 35, pages 5 – 18

DOI:10.1068/a3569

Estimating interregional economic impacts: an evaluation of nonsurvey, semisurvey, and full-survey methods

Jan Oosterhaven
Department of Spatial Economics, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands; e-mail: J.Oosterhaven@eco.rug.nl

Ed C van der Knijff¶, Gerard J Eding§
Province of Fryslân, Postbus 20120, 8900 HM Leeuwarden, The Netherlands
Received 5 March 2002; in revised form 20 July 2002

Abstract. Literature shows that nonsurvey input – output tables tend to produce regional multipliers with systematic upward biases. This paper explores the related, relatively uncharted territory of nonsurvey versus survey impact studies by means of a series of simulations. The base case is provided by a very detailed five region survey of both the forward and the backward impacts of the energy-distribution sector in the four northern provinces of the Netherlands. To deal adequately with the two-sided dependence between a firm or sector and a region, as opposed to using the traditional (gross) multipliers, we argue in favor of using a new multiplier concept: the *net* multiplier. Next, from examining alternative impact study methods ranging from quick and dirty, via semiextensive, to full-survey methods, we conclude that using even aggregate, first-order impact information is more important than using a very detailed survey-based input – output model.

1 Introduction

In the Netherlands, a strong tradition in constructing, updating, and using (inter)regional input – output (IO) tables has been built up. The development up to the 1970s can be summarized as running "from regional tables with only limited information used for primarily descriptive purposes towards ideal interregional tables mainly used for analytical purposes, such as estimates of economic impacts, experiments with programming models and building full forecasting models" (Oosterhaven, 1981, page 24).

As opposed to the tables constructed in the 1970s and opposed to most tables constructed in the USA (Brucker et al, 1990) and in Australia (see West, 1990), the Dutch (bi)regional tables of the 1980s are based mainly on surveys of *export coefficients* instead of the usual approach via import coefficients or regional purchase coefficients [RPCs (Stevens et al, 1989)]. This change in approach is the outcome of several surveys showing that firms have more and better data on the spatial destination of their outputs than on the spatial origin of their inputs. This new approach became standardized in the double-entry, biregional construction method [DEBRIOT (Boomsma and Oosterhaven, 1992)].

In the area of impact studies, the Dutch experience also produced results that are useful for a wider audience, especially where the estimation of forward impacts and export substitution is concerned (Oosterhaven, 1981; 1988). Richardson (1985) gives a summary of the wider known discussion of estimating the backward impacts of a new firm or sector by means of adding an extra IO row and column to avoid double-counting impacts, and changing the old **A**-matrix to account for import substitution (Hewings, 1971; Morrison, 1973; Tiebout, 1967). Still, there remains the relatively uncharted area of using survey versus nonsurvey data to estimate the first-order

¶ Present address: Province of Flevoland, Postbus 55, 8200 Lelystad, The Netherlands; e-mail: knijff@flevoland.nl

§ Present address: Statistics Netherlands, Postbus 4000, 2270 JM Voorburg, The Netherlands; e-mail: gedg@cbs.nl

impacts of a new firm or project. This issue is particularly important in practice, because resources are often insufficient for a full-fledged survey-based impact study. However, hardly any recent research can be found on this issue.

Spijker (1985) gives a flavor of the type of issues involved. In the context of a tourism impact study a mixed survey – nonsurvey type of single-region IO table was constructed from a survey into four tourist-specific sectors. The RPCs that related to the inputs of the tourist sectors were used to regionalize the national input coefficients that were assumed to hold for the other eighteen regional sectors. A complete table resulted from putting the four survey and the eighteen nonsurvey columns into one single table. The type-I multipliers (Miller and Blair, 1985) from this table were compared with those from a more detailed semisurvey biregional table (FNEI, 1984). This showed differences in the indirect part of the multipliers for the four tourist sectors of only 5 – 10%, whereas the differences in the indirect part of the eighteen other multipliers ran from -50% to $+65\%$, with outliers as large as $+180\%$, $+280\%$, and $+510\%$.

We believe that this result shows that both Bourque (1990) and Beemiller (1990) are right in their discussion in the *International Regional Science Review.* Bourque is right in his rejection of RIMS's nonsurvey alternative for the Washington State IO table, and Beemiller is right in his claim that combining direct information for the sectors of an impact study with a nonsurvey table produces sufficiently accurate estimates for most practical impact questions. But this assumes that full-survey data are available for the sector or project in question, and that the use of nonsurvey data is limited to the IO table.

Here we will investigate the opposite case, and concentrate on the size and type of errors made when a close-to-survey interregional IO table is combined with different types of nonsurvey data for the sector or project. To this aim, we first discuss the setup and the main results of a rather extensive survey-based impact study for the main energy distribution company in the east and the north of the Netherlands (EDON). Aside from its rather detailed base data, this study also provides for a new way to deal with the existence of forward impacts. To this aim it proposes to use a new *net* impact multiplier instead of the traditional (gross) multiplier. After this, the results of a series of nonsurvey estimations of the traditional backward impacts will be shown, ranging from a 'quick and dirty' method, via 'quick and nondirty', 'semiquick', and 'semiextensive' methods. All will be compared with the results of the actually used full-survey method. The implications of this research into the accuracy of nonsurvey methods for doing impact studies will be drawn in the concluding section.

2 The EDON impact study

Practically all impact studies concentrate on the 'effects of' or the 'importance of' a certain branch of industry or a specific project. Seldom is the question asked or answered whether, from a model of view, this sector or project might be considered as being exogenous or endogenous. Implicitly, most impact studies assume that the sector or project is exogenous to the region, and continue the estimation of the impacts accordingly. However, when part of the sector or project is endogenous to the region, this will lead to double-counting impacts, as will be shown.

More generally, this implicit assumption leads to an unbalanced presentation of the importance of the sector or project. In fact, if the traditional type of impact study were done for all the sectors of an economy and the estimated impacts were added, one would have to conclude that the economy was two to three times larger than its actual size. Or to be more precise: the 'estimated' size of the economy would be equal to its actual size multiplied by the weighted average size of its sectoral multipliers (Oosterhaven and Stelder, 2002).

A more balanced presentation of the importance of a sector for a certain region, therefore, should start with a discussion of the dependence of the project on the economy of the region. This is precisely the first question answered in the EDON impact study (Cras et al, 1995). As a consequence of this different beginning, this particular impact study emphasizes the *mutual* dependency of the energy-distribution company and its regional economy, instead of the usual one-sided story of the region's dependence on the sector or project.

Besides this, there is a second, even more general, point that is often neglected in impact studies. It relates to the type of research question that is being answered. In principle, two basically different types of questions may be posed, each requiring a different modeling approach.

First, one might want to know which existing activities, and how far existing activities, are or would be dependent on the *existence* of a certain sector or project. When using an (inter)regional IO model, this question requires the use of *average* coefficients, along with the use of type-II income or employment multipliers (Batey, 1985; Miller and Blair, 1985). This is the traditional approach that is also used in the present study. Quite often, however, the results from such an impact study are (mis)used to answer the next, quite different, kind of question.

For, second, one might want to estimate the *change* in economic activity that is the result of a change in the sector or that is the result of the *execution* of the project under study. For example, when the sector increases its activities, supplying firms might need only a few more employees to satisfy the increase in intermediate demand. Thus, *marginal* employment coefficients are needed instead of average ones. Moreover, not all employees may be immigrants from different regions, which is the implicit assumption when type-II multipliers are used. Some may be drawn from other sectors in the same region, and this requires a vacancy-chain approach (Van Dijk and Oosterhaven, 1986). Others may be regionally unemployed or otherwise nonactive and consequently lose their entitlement to different types of benefits, and this requires the use of type-III or type IV multipliers (Batey, 1985; Oosterhaven and Dewhurst, 1990).

The EDON impact study answers the first kind of question in three steps. (1) An extensive survey of the company's sales, purchases, and personnel costs was made. (2) A five-region impact model was constructed for the provinces of Groningen, Friesland, Drenthe, and Overijssel, which constitute the two main regions of operation of the EDON company. The fifth region comprises the aggregate of the remaining eight, more southern, provinces of the Netherlands. (3) The results of the first two steps were combined to estimate the economic importance of EDON for the regions in question. This third step started with the survey estimate of the *direct and forward* effects. Then the *backward* effects were determined with the impact model. Here, we will briefly discuss both the model and the main empirical results of this study.

The interregional *impact model* we used estimates the importance of any sector with respect to two impact variables, namely regional indirect income (that is, gross value added at market prices, v_r) and regional indirect employment (w_r). This is done by multiplying indirect production per sector, per region (x_{ir}) by the average income coefficients (c_{ir}) and the average employment coefficients (e_{ir}) for the forty-seven sectors (i) and the five regions (r) of the model:

$$v_r = \sum_i c_{ir} x_{ir}, \qquad w_r = \sum_i e_{ir} x_{ir}, \qquad \forall r. \tag{1}$$

The EDON survey data relate to 1994, whereas the average employment coefficients (e_{ir}) relate to 1990. Hence, the latter needed to be updated for 1990–94 price and real labor-productivity development, per sector, per region (for details see Cras et al, 1996).

The indirect production (x_{ir}) is calculated from the total of the demand of the EDON branches located in region s and the demand of its employees working in region s ($d^{\text{edon}}_{ir,s}$), by means of the following equations:

$$x_{ir} = \sum_j \sum_s a_{ir,js} x_{js} + \sum_{s'} p_{ir,s'} \left(\sum_j \sum_s b_{s',js} x_{js} \right) + \sum_s d^{\text{edon}}_{ir,s}, \qquad \forall i \text{ and } r. \quad (2)$$

The first term defines the endogenous intermediate demand for product i from region r by sectors j in regions s. The second term defines the endogenous consumption demand for product i from region r by workers living in regions s', working in sectors j in regions s. The third term defines the total demand from EDON and its workers from the different regions s.

In equation (2) the average intermediate input coefficients ($a_{ir,js}$) are derived from a five-region IO table for the Netherlands for 1990. This table was constructed from existing biregional IO tables for the provinces of Groningen, Friesland, Drenthe, and Overijssel for 1990 (Eding et al, 1995a). The four 2×2 tables were telescoped into one single 5×5 table, by using survey data for the total trade per sector between each of the three northern provinces from Eding et al (1995b), and using the nonsurvey gravity method from Oosterhaven (1981, appendix) to fit in the fourth province of Overijssel (for details, see Cras et al, 1995).

In equation (2) the average consumption package coefficients ($p_{ir,s'}$) indicate which part of the consumption expenditures of employees and dependents *living* in region s', is spent on goods and services from sector i in region r. They are directly derived from the five-region IO table. The average total consumption ($b_{s',js}$) indicates the total amount of consumption expenditures of workers *living* in region s' and *working* in sector j in region s, per unit of production of sector j in region s. These coefficients result from a multiplication of interregional commuting coefficients, ratios of total consumption to gross labor income, and unit gross labor income coefficients per sector, per region.

Most impact studies use equations (1)–(2), or simplifications thereof, without further consideration. However, when the impacts of the total production of a sector are estimated, and not just the impacts of its exogenous final output, equation (2) produces an overestimation of the real impact of the sector. This is because x_{ir} in equation (2) includes (some of the) intermediate sales and endogenous consumption sales of the sector, but these outputs are already included in the total production that was treated as exogenous in $d^{\text{edon}}_{ir,s}$. To prevent such *double counting*, the rows with coefficients relating to the intermediate and endogenous consumption output of the sector are set equal to 0 (Oosterhaven, 1981, chapter 8).(1) Unfortunately, this is often forgotten.

With this provision, the adapted equations (1) and (2) are easily put into matrix algebra, which gives the following solution for the impact model:

$$\boldsymbol{v} = \mathbf{C}(\mathbf{I} - \mathbf{A}^{\text{edon}} - \mathbf{P}^{\text{edon}}\mathbf{B})^{-1}\boldsymbol{d}^{\text{edon}} \equiv \mathbf{C}\mathbf{L}^{\text{edon}}\boldsymbol{d}^{\text{edon}}, \quad (3a)$$

$$\boldsymbol{w} = \mathbf{E}(\mathbf{I} - \mathbf{A}^{\text{edon}} - \mathbf{P}^{\text{edon}}\mathbf{B})^{-1}\boldsymbol{d}^{\text{edon}} \equiv \mathbf{E}\mathbf{L}^{\text{edon}}\boldsymbol{d}^{\text{edon}}, \quad (3b)$$

in which

$\boldsymbol{v}$, $\boldsymbol{w}$ are 235-vectors ($235 = 47$ sectors $\times 5$ regions) with indirect income and indirect employment,

C, **E** are diagonal 235×235-matrices with average income and average employment coefficients,

(1) Alternatively, one may estimate the impacts of *exogenous* final output $\boldsymbol{f}^{\text{exo}}$, instead of the impacts of *total* $\boldsymbol{d}^{\text{edon}}$, and then one may use the regular $\boldsymbol{w} = \mathbf{E}(\mathbf{I} - \mathbf{A} - \mathbf{PB})^{-1}\boldsymbol{f}^{\text{exo}}$ instead of equation (3b).

$\mathbf{A}^{\text{edon}}$ is a 235×235-matrix with average intermediate input coefficients, with the EDON rows equal to 0,
$\mathbf{P}^{\text{edon}}$ is a 235×5-matrix with average consumption package coefficients for the five regions of residence, with the EDON rows equal 0,
$\mathbf{B}$ is a 5×235-matrix with average total 'consumption from labor income' coefficients for the five regions of residence,
$\boldsymbol{d}^{\text{edon}}$ is a 235-vector with the total demand of EDON companies and EDON employees, as estimated by the survey.

Next, the *importance* of EDON for the economies of the regions distinguished in equations (3) is determined by estimating five types of impacts. Table 1 shows the results for the employment impacts for the three main regions.

First, the dependence of EDON *on* the economies of the four northern provinces is determined. As EDON is the main regional distribution company for natural gas and electricity, almost all its output is sold in these four regions. Within the Netherlands, price differentials for *natural gas* are less than 10%, and the large industrial users are supplied directly by the national gas distribution company (Gasunie). Hence, for its natural gas sales, EDON is entirely dependent on the income growth of the region's inhabitants and the output growth of the region's firms. It has no influence (through its sales of natural gas) over the size of both. Hence, the company's forward natural gas linkages do not produce any forward employment impacts. Interregional wage differences and other location factors are more important. Causality does not run forward, but runs the other way (for an overview, Oosterhaven, 1996).

The same holds for most of EDON's sales of *electricity*. But here the situation is a little more complicated. EDON has individual contracts with the forty largest industrial users of electricity, and with the other large users regional price differences of up to 30% are noted. Hence, EDON potentially might attract firms to the region by means of favorable contracts. A detailed analysis of its individual energy-intensive users showed that three of them have contracts of such a nature that their presence in the region might be explained by the these contracts. The firms produce aluminium (Aldel) and magnesium (ESD) by means of electrolysis, and offer research services (Windtunnel). All three are highly dependent on the electricity price and delivery conditions. In fact, these three firms alone purchase 5% of all EDON's output.

It was decided to consider only this 5% of all forward linkages to cause *forward* income and employment *effects*. These despite the fact that it might almost as easily be argued that other locational conditions—such as labor and the local availability of raw magnesium salts, or the availability of cheap industrial sites and absence of congestion—are dominant in the location/production decisions of these three firms (for the transport sector, see Oosterhaven et al, 2001). When these three firms are

Table 1. Aggregate employment importance of EDON, number of jobs in 1994 (source: Cras et al, 1995, and own calculations).

Impact	Type of effect	Northern Netherlands	Province of Overijssel	Rest of Netherlands	Total Netherlands
Sales	forward and related backward	1 530	30	540	2 100
Size of EDON	direct	1 560	1 730	0	3 290
Energy purchases	backward	610	220	720	1 550
Other purchases	backward	1 000	950	2 300	4 250
Wages and salaries	backward	420	280	340	1 040
Total gross impact		5 120	3 210	3 900	12 230
Total net impact		1 710	190	710	2 610

considered to depend solely on EDON, then this of course also applies to their backward linkages. The latter are estimated by applying the basic metal industry's multipliers from the five-region impact model to the production levels of the three firms.[2] In this way a national total of 2100 jobs is estimated to be dependent forwardly on the sales of EDON (see table 1).

Second, the *direct importance* of EDON was determined by the survey. Note that 95% of this direct importance is endogenous, that is, dependent on the size and growth of the regional economy. In employment terms EDON accounts for 3290 jobs, more than half of which are found in the province of Overijssel, the region where the new headquarters of the two constituent parts are located.

Third, the direct *purchases of energy* by sector and region of origin are determined by means of the survey. The backward input of these purchases is estimated from equation (3b), and amounts to 1550 jobs, nationally. The relatively large size of this employment impact surprised us a little, in view of the capital-intensive character of most producers of primary energy (natural gas and electricity), but no obvious errors could be found.

Fourth, the direct *other purchases* consist of 50% purchases for *intermediate use* and 50% purchases for *investment purposes*. In view of the volatility of the investment expenditures, a time series was used to check the constancy of the EDON survey data for 1994, which was found to be a fairly average year. This second backward employment effect in table 1 is based on the survey's sectoral and regional origin data for the total of all other purchases. It amounts to 4250 jobs nationally, which is a large figure and follows directly from the capital investments needed to maintain and extend the energy-distribution production process.

Finally, the *wages and salaries* of the 3290 employees of EDON lead to consumption expenditure, and to a related backward employment of 1040 jobs in the Netherlands as a whole.

Two numbers may summarize the result of the analysis. First, the traditional employment multiplier, which we will label the *gross* employment multiplier, and, second, a new concept, the *net* employment multiplier, which equals the gross multiplier corrected for the endogeneity of the direct and related backward impacts of the company or project in question (Oosterhaven and Stelder, 2002).

Thus, the *gross regional* employment multiplier that follows from table 1 is as large as 2.5 [(5120 + 3210)/3290] and the *gross national* employment multiplier amounts to as much as 3.7 (12230/3290). Naturally, being a rather capital-intensive industry with relatively little direct employment of its own, energy distribution is bound to have a relatively large multiplier whenever both forward and investment effects are included. When one excludes the forward and investment impacts the gross regional employment multiplier comes down to 1.6 and the gross national multiplier comes down to 2.1, which are both numbers that are more commonly found in the literature.

Nevertheless, all four numbers suggest that energy distribution is important for both regional and national employment. When we look at the way in which EDON depends on other activities, the above numbers can be put into perspective. Energy distribution is primarily a service industry, with 95% of its activity fluctuating with the welfare of the region. Only its forward employment (2100), the related 5% of its own employment (165), and the related 5% of its backward employment (345) may be considered exogenous to the economic size of the region it serves. Consequently, the *net national* employment multiplier equals 0.8 (2610/3290) and the *net*

(2) In this case, the rows with intermediate input coefficients and consumption package coefficients for the basic metal industry as well as those of EDON had to be set equal to 0 (compare Oosterhaven, 1981, chapter 8; 1988, figure 2).

regional employment multiplier is even smaller and equals only 0.6 (1900/3290). Both numbers are smaller than 1.0, indicating that the net causality runs from the region to the sector and not the other way.

Next, we will turn to the question about the size of the errors made when the backward impacts are estimated by means of nonsurvey methods instead of using the actual, full survey into the company's records.

3 Nonsurvey methods for estimating backward impacts

To keep the analysis simple, the five-region impact model used in this section will exclude the consumption effect, both exogenously and endogenously. Hence, in equation (3), $\boldsymbol{d}^{\text{edon}}$ will be exclusive of the consumption expenditures of the EDON workers, and $\mathbf{L}^{\text{edon}}$ will be exclusive of $\mathbf{P}^{\text{edon}}\mathbf{B}$. Thus, type-I multipliers will be used and compared instead of type-II multipliers. More precisely, because (IO) models are used to estimate the indirect impacts and not to estimate the direct impacts, only the indirect effects will be compared.[3]

Impact studies can be done, and are done, using many different methods. These range from quick and dirty methods, where only information on total industrial output is used in combination with a set of aggregate multipliers, towards full-survey methods, where a maximum amount of direct survey information is used in combination with a detailed input–output model, as in the EDON case. In this section various less than full-survey methods will be explored and the matching mathematical formulas will be specified. The EDON data are used for testing purposes and for each method the estimate of total *indirect national* employment will be presented. In subsection 3.5 these aggregate results will be compared further, both in terms of the regional composition and in terms of the sectoral composition of the aggregate national impact.

3.1 Quick and dirty: the standard multiplier application

The first method represents the simplest way to perform an impact study. Because of the minimal information needed, the approach is often found in quick commercial studies. The only requirements are: information on the (regional) production of the company or project and an aggregate (regional) multiplier for a more or less comparable existing sector. In the EDON case study, total output (x^{edon}) and the geographical division of employment over the four study regions ($\alpha_r^{\text{edon}} = w_r^{\text{edon}}/w^{\text{edon}}$) is simply taken from the annual report. This information is sufficient to derive an approximation of EDON's regional production. Second, the public utility sector is taken to be representative of the energy-distribution sector. With these data and this assumption it is possible to fill the five-region $\boldsymbol{d}^{\text{edon}}$ vector in equation (3) as follows:

$$\begin{cases} d_{ir}^{\text{edon}} = \alpha_r^{\text{edon}} x^{\text{edon}}, & \text{for } i = \text{public utilities,} \\ d_{ir}^{\text{edon}} = 0, & \text{for } i \neq \text{public utilities.} \end{cases} \tag{4}$$

To estimate the employment impacts of EDON production quickly, the nonzero cells of $\boldsymbol{d}^{\text{edon}}$ are simply multiplied by the standard aggregate regional or national employment multiplier of the public utility sector at large. Thus, the *indirect* employment impact per region is derived from

$$\boldsymbol{w} = \mathbf{E}(\mathbf{L} - \mathbf{I})\boldsymbol{d}^{\text{edon}} . \tag{5}$$

In equation (5), the unity matrix $\mathbf{I}$ is subtracted from the Leontief inverse $\mathbf{L}$ in order to exclude direct employment, which is already known to be 3290 nationally (table 1).

(3) The misleading standard approach is to compare multipliers and the percentage errors therein. Consequently, the standard type of comparison systematically undervalues the estimation errors of the model. This bad practice may be one of the reasons for the continued faith in the IO model.

The result of substituting equation (4) in equation (5) is an estimated total indirect national employment of 4977 full-time jobs. Note that this estimate may not be compared with those in table 1, which relate to wider impacts estimated with a more extended impact model.

3.2 Quick and nondirty: excluding double counting

The above method is labeled *dirty* because the impact on regional employment is estimated with a systematic upward bias. If we multiply the Leontief inverse by total output, instead of by *exogenous* final demand, the indirect impacts also include the production of the sector (EDON) that results from energy purchases of other sectors, which are already included in the direct effect. To exclude this type of double counting, the Leontief inverse in equation (5) needs to be replaced with the Leontief inverse based on an **A**-matrix with zeros on the public utility row ($\mathbf{L}^{\text{edon}}$), as explained in section 2.

Thus in a *quick but nondirty* method equation (5) will be replaced by

$$\boldsymbol{w} = \mathbf{E}(\mathbf{L}^{\text{edon}} - \mathbf{I})\boldsymbol{d}^{\text{edon}} . \tag{6}$$

When equation (4) is substituted into equation (6), total indirect employment is estimated at 4670 jobs as compared with the earlier estimate of 4977 jobs. Thus, in the EDON case, dirty implies an upward bias of 7% of the indirect impact. The size of the bias will largely depend on the relative size of the regional or national (endogenous) intermediate sales in total output of the sector or company that is the subject of analysis. The larger the relative size of the intermediate sales of the sector or company, the more important the correction for double counting (Oosterhaven and Stelder, 2002).

3.3 Semiquick: using aggregate survey data

The above quick methods assume the structure of the columns of the five-region IO table (RIOT) for the public utility sector to be the same as those of the regional branches of EDON.(4) This standard assumption is, of course, incorrect. Here we will investigate the possibilities of reducing the estimation errors by means of relatively easily accessible aggregate survey data.

First, the effect of using survey data on total purchases and total value added will be studied. Second, the company's data on total energy purchases versus all other purchases will be used. Third, as energy inputs appear to be rather important, it seems worthwhile substituting the detailed company data on the sectoral and regional origin of its energy purchases for the origins implicit in the five-region IO table. Recovering this last type of data from the company's record was not really *semiquick*, but took far less time than the further subdivision of the other purchases, which will be discussed in the next paragraph.

The result of replacing aggregate RIOT data with aggregate EDON data will be discussed in a stepwise fashion, so that the influence of each more detailed type of survey information can be identified sequentially.

3.3.1 *Survey data on total intermediate input*

The first step is based on generally available company information (for instance, from annual reports), in addition to the previously used company information on total output and regional direct employment. It concerns the aggregate share of value added (mainly wage costs and profits) and the complement aggregate share of all intermediate purchases in total output ($a_{\bullet\bullet,j}^{\text{edon}}$). The further subdivision of this last total for the four regional public utility sectors will be made by means of the data from the

(4) This follows from the fact that in equations (5) and (6): $\mathbf{L} - \mathbf{I} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{A}$,

five-region table ($a^{\text{riot}}_{ir,\,js}$). As a formula this method implies

$$d^{\text{edon}}_{ir,\,s} = \left(\frac{a^{\text{riot}}_{ir,\,js}}{a^{\text{riot}}_{\bullet\bullet,\,js}}\right) a^{\text{edon}}_{\bullet\bullet,\,j}\,\alpha^{\text{edon}}_{s}\,x^{\text{edon}}\,, \tag{7}$$

where *js* indicates the public utility sector in region *s*, and • indicates summation over the index concerned.

The survey share of total intermediate purchases ($a^{\text{edon}}_{\bullet\bullet,\,j}$) appears to be relatively large (0.72) compared with the (regionally weighted) share from the IO table ($a^{\text{riot}}_{\bullet\bullet,\,js}$) of 0.51. Because $\boldsymbol{d}^{\text{edon}}$ in equation (7) now represents intermediate demand, instead of total output as in equation (4), there is no more need to correct for direct employment by means of subtracting the unity matrix **I**. Therefore equation (6) is replaced by

$$\boldsymbol{w} = \mathbf{E}(\mathbf{I} - \mathbf{A}^{\text{edon}})^{-1}\boldsymbol{d}^{\text{edon}} \equiv \mathbf{E}\mathbf{L}^{\text{edon}}\boldsymbol{d}^{\text{edon}}\,. \tag{8}$$

Substituting equation (7) in equation (8) leads to a new indirect impact estimate of 6564 jobs, instead of the previous 4670. Replacing the nonsurvey RIOT approximation of the total intermediate purchases of EDON with the survey value thus results in an upward correction of indirect employment by as much as 41%.

3.3.2 *Survey subdivision between energy and other intermediate inputs*

A further refinement is the replacement of the RIOT distribution of total intermediate purchases over energy purchases and other purchases in equation (7) with easily accessible data from the EDON survey. As a formula:

$$d^{\text{edon}}_{ir,\,s} = \left[\left(\frac{a^{\text{riot}}_{ir\subset er,\,js}}{a^{\text{riot}}_{e\bullet,\,js}}\right) a^{\text{edon}}_{e\bullet,\,j} + \left(\frac{a^{\text{riot}}_{ir\subset or,\,js}}{a^{\text{riot}}_{o\bullet,\,js}}\right) a^{\text{edon}}_{o\bullet,\,j}\right]\alpha^{\text{edon}}_{s}\,x^{\text{edon}}\,, \tag{9}$$

where *js* indicates public utilities in region *s*, *e* are energy inputs, *o* are other, nonenergy inputs, and • indicates summation.

The energy inputs appear to have a larger share (0.89) in the total intermediate inputs of EDON than the (regionally weighted) RIOT average for the utility sector (0.82) suggests. Consequently, the other purchases have a smaller share, that is, 0.11 instead of the RIOT average of 0.18 from the five-region table. This further addition of survey data, that is, substituting equation (9) into equation (8), leads to a reduction of 17% of the estimated indirect employment impact of 6564 jobs to 5426 jobs, only 1895 of which are found in the nonenergy sectors. The reason for the downward correction is that producing energy inputs requires less labor than producing other inputs of an equal amount of euros.

3.3.3 *Semiextensive: survey data on energy inputs*

Because energy inputs represent the single most important type of intermediate input of EDON, a survey-based desegregation of these inputs seems the most logical next step in improving the impact estimate. In equation (9), energy inputs are treated as the average product of the utility sector, that is, as the average of the production and distribution of water, gas, and electricity. The annual report of EDON gives more information about the type of energy that is actually used, about the actual location of the (few) primary energy suppliers, and about the size of these purchases. Hence, the nonsurvey RIOT division of energy inputs in the first part of equation (9) may be relatively easily replaced with full-survey information on the precise sector of origin ($a^{\text{edon}}_{i\bullet\subset e\bullet,\,j}$) and the precise region of origin ($t^{\text{edon}}_{ir\subset er,\,j}$):

$$d^{\text{edon}}_{ir\subset er,\,s} = t^{\text{edon}}_{ir\subset er,\,j}\,a^{\text{edon}}_{i\bullet\subset e\bullet,\,j}\,\alpha^{\text{edon}}_{s}\,x^{\text{edon}}, \qquad \forall i \subset \text{energy}\,. \tag{10}$$

This additional information proves to be very important indeed. Replacing the first term in equation (9) with equation (10) leads to a 42% reduction of the last estimate of 5426 jobs to 3145 jobs.

This sequence of adding increasingly more survey data shows that each separate step will not always lead to an improvement of the estimate (table 2). Thus, determining when to stop adding survey data at the intermediate level is a tricky affair. This is an important reminder of the uneasy fact that the probability of compensating errors becomes smaller when more survey data are added.

Table 2. Percentage errors in national employment compared with the full-survey type-I impact estimate.

Type of estimation model	Total indirect impact (%)	MAPE[a] of impacts of nonenergy inputs (%)	
		over 11 sectors	over 5 regions
Quick and dirty	+60		
Quick and nondirty	+50		
Semiquick: intermediate total	+111		
Semiquick: energy versus other inputs	+76		
Semiextensive: survey data on energy	+1.2	53	129
Extensive: plus survey industry codes	+0.5	1	106
Extensive: plus survey postal codes	+1.3	63	7
Survey data: first-order impacts only	−17	37	24

[a] MAPE mean absolute percentage error.

3.4 Extensive: survey data on nonenergy intermediate inputs

Finding survey data on energy inputs is relatively easy when an energy-distribution company is being considered. Comparable improvements may be made in other impact studies when a few major inputs can be surveyed at relatively low cost. Surveying the host of other minor inputs will involve much more time. With a large company like EDON, intermediate inputs are diverse, ranging from electrical equipment to cleaning services, and are purchased from a wide variety of suppliers in different regions. In the extensive, actual survey, all these purchases have been assigned to a producing industry and a producing region. In particular, deliveries by retailers and wholesalers required an extra survey step to identify correctly the ultimate sector and region of origin (see Boomsma and Oosterhaven, 1992).

In this subsection we will discuss the importance of such further detail. First, we discuss the effect of using 'exact' information on the sectoral origin of the other purchases and, second, the effect of having 'exact' information on the regional origin. In the EDON case both types of data were recovered from the companies' internal purchase administration simultaneously, by assigning regions to the postal codes of the suppliers and by assigning industry codes to the supplying companies. The industry codes per company were obtained from external statistical sources. In the following, in order to highlight the possible differences, these two types of data will be used separately, despite their simultaneous derivation from EDON's purchase administration.

3.4.1 *Survey data on sectoral origin*

First, the precise industry-code distribution of the nonenergy suppliers of EDON (a^{edon}) is combined with the trade coefficients of the public utility sectors (j in region s) from the five-region table (t^{riot}). Hence, the nonenergy cells in the $\boldsymbol{d}^{\text{edon}}$ vector are calculated from

$$d^{\text{edon}}_{ir \subset or,\, s} = t^{\text{riot}}_{ir \subset or,\, js}\, a^{\text{edon}}_{i\bullet \subset o\bullet}\, \alpha^{\text{edon}}_{s} x^{\text{edon}}, \qquad \forall i \subset \text{nonenergy sectors}. \tag{11}$$

Substituting equation (11) directly into equation (8) gives an estimate of 1875 jobs as opposed to 1895 when the survey information on the industry codes is not used. Hence, the new estimate is quite close to the old estimate of the impact of the purchases of other, nonenergy inputs. Though the national total employment impact hardly changes, the sectoral distribution of that total is now considerably better, whereas the regional distribution is still pretty bad (table 2).

3.4.2 *Survey data on regional origin*

Next, the opposite combination of survey and nonsurvey data is investigated, that is, the precise postal codes from the survey (t^{edon}) are combined with the technical coefficients for the public utility sector (j in region s) from the five-region table (a^{riot}), as follows:

$$d^{\text{edon}}_{ir \subset or,\, s} = t^{\text{edon}}_{ir \subset or}\, a^{\text{riot}}_{i\bullet \subset o\bullet,\, js}\, \alpha^{\text{edon}}_{s} x^{\text{edon}}, \qquad \forall i \subset \text{nonenergy sectors}. \tag{12}$$

Substituting equation (12) directly into equation (8) gives an alternative estimate of the impact of the nonenergy inputs. This estimate appears to be even closer to the previous one: 1898 jobs instead of 1895. In terms of improvement of the estimated national total, the second extensive method does not pay off either. However, it does lead to a considerably better estimation of the regional distribution of these impacts (table 2).

The reason is that the energy-distribution company EDON purchases far more inputs from the remaining part of the Netherlands than does the regional utility sector at large. This difference results in smaller shares for the employment impacts in Groningen and, especially, Friesland, and into a larger share for the rest of the Netherlands (table 3). A relatively accurate impression of the regional distribution can also be achieved by an examination of the first-order survey-based impacts (table 3). Thus: if a reliable estimate of the regional distribution of impacts is the main aim of analysis, good survey data on the regional origin of the purchases appear to be more helpful than setting up a full-survey IO model.

Table 3. Regional distribution (%) of indirect employment impacts of nonenergy purchases.

Region	Using RIOT[a] trade data	Using EDON trade data	First-order survey impact
Groningen	17	13	14
Friesland	13	3	2
Drenthe	8	10	11
Overijssel	23	24	26
Rest of the Netherlands	39	51	46

[a] RIOT five-region input–output table.

3.5 Conclusion: comparison with the full-survey estimate

Assuming that the impact model that uses all available survey information gives the most accurate description of reality, the preceding nonsurvey approximations can best be judged on accuracy in comparison with the full-survey model. In table 2 this is done with respect to the total backward employment impact, and with respect to the sectoral and the regional distribution of the impact of the nonenergy inputs.

Clearly, all *quick* methods lead to a considerable overestimation of the total indirect impact. This result cannot be generalized, however. If other subsectors of the public utilities sector are the subject of quick methods of impact assessment, a compensating underestimation may well be the result, assuming that the five-region table as such is reliable. It is also very important to note that the *semiquick* methods (+76% and +111%) are not necessarily better than the *quick* methods (+50% and +60%).

Another feature is that estimating only first-order indirect impacts of course leads to underestimation. But in the EDON case this underestimation proved to be only −17%, which is a smaller error than the considerable overestimations of the various quick methods (+50% to +111%). Thus, it may be more important to search for even limited survey information than to invest in building a full-survey IO table and (mis)-use the sectors from that table as nonsurvey proxies for the actual sector or company under consideration. However, when type-II multipliers are used as in equation (3), the under-estimation of using only first-order impacts will be larger, as the (higher order) indirect impacts will become larger.

The importance of survey information is confirmed by the pattern of estimation errors of the *extensive* methods. The use of survey data on only the energy inputs already leads to a very accurate estimate of the *total* national impact (+1.2%). The estimated sectoral and regional distribution of this total, however, is quite wide of the mark. When the postal code origins for the nonenergy inputs are added the mean percentage error (MAPE) of the regional shares reduces from 129% to only 7%. The same holds for the industry codes. When survey industry code origins are used the MAPE for the sectoral distribution reduces from 53% to only 1%. Thus, the survey of the purchases of the company or sector should be concentrated on the (sectoral or regional) type of impact information one is most interested in.

In interpreting these results further, it should be realized that not all conclusions may be generalized. First, more impact simulation studies like this one have to be done. Nevertheless, some provisional general conclusions may be drawn:
(1) impact assessments of individual activities with aggregate (interregional) multipliers alone, is a most risky type of analysis,
(2) assuming homogeneity within sectors is rather hazardous, unless extremely detailed IO data are used,
(3) aggregate information on the most important items in the cost structure of the activity most probably will improve the accuracy of the indirect impact considerably,
(4) full-survey data seem to be useful only when one is interested in the precise sectoral or regional distribution of impacts.
For a discussion on the related issue of partitive versus holistic accuracy of IO tables, see Jensen (1980), and for a discussion on the related issue of analytically important cells of input coefficient matrices, see Sonis and Hewings (1992).

4 Conclusion

First, this paper shows that it is extremely important to ask the right questions before starting an impact study. Only by examining the *mutual* dependence between the economic activity at hand and the rest of the regional economy, may one put the so-called impacts of that particular activity in its proper perspective. Therefore, the traditional impact multiplier may be better labeled the *gross* impact multiplier. In our case, the EDON company appeared to be more dependent on the rest of the regional economy than vice versa. Maximally 5% of its output could be considered to induce forward linkages, the rest had to be considered as being dependent on regional demand, and none was dependent on exogenous demand. To capture the net effect of the two opposing dependencies a new concept is needed: the *net* impact multiplier.

In the case of the EDON company the net regional employment multiplier, in fact, amounted to 0.6, whereas the gross multiplier was a misleadingly large 2.5.

Second, it is important to separate average and marginal linkages. Only when the simple 'dependency question' is asked, is the relatively simple assessment of the average linkages by means of type-II multipliers acceptable. If EDON had asked us about the regional and national impact of a reduction or increase in its production, in whatever shape, a much more complicated analysis would have been required. This would have involved looking at the potential for substitution between different energy inputs by other firms, looking at marginal impacts using employment elasticities, and considering changes in unemployment benefits, etc.

A third important conclusion regards the estimation errors of nonsurvey impact estimates. These errors relate both to the total impact and to its distribution over regions and sectors. In this particular case, the total national impact could have been assessed accurately with an analysis of aggregate company data and its most important inputs and outputs. The respective sectoral or regional distribution of the effects, however, could only be assessed accurately with an in-depth analysis of the respective sectoral or regional origin codes of all inputs and outputs.

In general, estimation errors can roughly be divided into (1) the lack of IO table precision, (2) the lack of homogeneity within the relevant sector of the IO table, and (3) the lack of survey data on the inputs and outputs of the company or sector under consideration. Consequently, there exists a complicated trade-off between each of these three types of lack of data. Having a full-survey IO table probably implies that less information is needed when doing an impact study, assuming that the level of detail of the IO table is sufficient to prevent large homogeneity errors. Having a low-cost nonsurvey table, on the other hand, will definitely require the collection of more detailed information on the activity as a compensation.

The EDON impact study shows that survey information on a company or sector might well be more important than having a full-survey IO table. Thus, the present observations confirm Beemiller's claim, quoted in the introduction, and lead to the conclusion that for the purpose of a single impact assessment it is best to invest in direct information on the particular activity. When a series of assessments are expected, survey information on the economy as a whole becomes relatively more important.

Acknowledgements. The authors thank N.V. E.D.O.N., presently a part of the N.V. Essent, for providing extremely detailed and confidential company data and for their openmindedness as regards our conclusions about their dependence on the rest of the regional economy, when they commissioned the original research (Cras et al, 1995). We also thank Jeroen Cras and Dirk Stelder for their contribution to the original impact study. The present extension of this study was first presented as "Regional Economic Impacts: Non-, semi- and full-survey methods applied to the energy distribution sector in the Northern Netherlands" by Van der Knijff at the 37th European Congress of the RSAI in Rome, August 1997, and was presented in a revised version at the 12th Conference of the IIOA in New York, May 1998. The authors thank the participants of both congresses, Geoffrey J D Hewings, and the anonymous referees for constructive comments.

References

Batey P W J, 1985, "Input – output models for regional demographic – economic analysis: some structural comparisons" *Environment and Planning A* **17** 73 – 99

Beemiller R M, 1990, "Improving accuracy by combining primary data with RIMS: comment on Bourque" *International Regional Science Review* **13** 99 – 101

Boomsma P, Oosterhaven J, 1992, "A double-entry method for the construction of bi-regional input – output tables" *Journal of Regional Science* **32** 269 – 284

Bourque P J, 1990, "Regional multipliers: WAIO vs. RIMS" *International Regional Science Review* **13** 87 – 98

Brucker S M, Hastings S E, Latham III W R, 1990, "The variation of estimated impacts from five regional input – output models" *International Regional Science Review* **13** 119 – 139

Cras J H A, Eding G J, van der Knijff E C, Oosterhaven J, 1995 *EDON en de Regio: Een sterke tweezijdige relatie* [EDON and the region: a strong bilateral relation] Stichting REG, University of Groningen, Groningen
Eding G J, Stelder T M, Vos E, Oosterhaven J, 1995a *Bi-regionale Interactie: Nieuwe input – output tabellen voor Groningen, Friesland, Drenthe en Overijssel voor 1990* [Biregional interaction: new input – output tables for Groningen, Friesland, Drenthe and Overijssel for 1990] REG-Publication 7, University of Groningen, Groningen
Eding G J, Stelder T M, Oosterhaven J, 1995b *Structuur en Interactie: Een interregionale structuurschets van Groningen, Friesland en Drenthe voor 1990* [Structure and interaction: the interregional economic structure of Groningen, Friesland and Drenthe for 1990] REG-Publication 10, University of Groningen, Groningen
FNEI, 1984 *De Interregionale Input – Output Tabel voor Drenthe-overig Nederland 1975* [The interregional input – output table for Drenthe and the rest of the Netherlands 1975] (Federatie van Noordelijke Economische Instituten, Assen)
Hewings G J D, 1971, "Regional input – output models in the UK: some problems and prospects for the use of nonsurvey methods" *Regional Studies* **5** 11 – 22
Jensen R C, 1980, "The concept of accuracy in regional input – output models" *International Regional Science Review* **5** 139 – 154
Miller R E, Blair P D, 1985 *Input – Output Analysis: Foundations and Extensions* (Prentice-Hall, Englewood Cliffs, NJ)
Morrison W I, 1973, "Input – output and the firm: some comments" *Regional Studies* **7** 253 – 256
Oosterhaven J, 1981 *Interregional Input – Output Analysis and Dutch Regional Policy Problems* (Gower, Aldershot, Hants)
Oosterhaven J, 1988, "On the plausibility of the supply-driven input – output model" *Journal of Regional Science* **28** 203 – 217
Oosterhaven J, 1996, "Leontief versus Ghoshian price and quantity models" *Southern Economic Journal* **62** 750 – 759
Oosterhaven J, Dewhurst J H L, 1990, "A prototype demo-economic model with an application to Queensland" *International Regional Science Review* **13** 51 – 64
Oosterhaven J, Stelder T M, 2002, "Net multipliers avoid exaggerating impacts: with a bi-regional illustration for the Dutch transportation sector" *Journal of Regional Science* **42** 533 – 543
Oosterhaven J, Eding G J, Stelder T M, 2001, "Clusters, linkages and regional spillovers: methodology and policy implications for the two Dutch mainports and the rural north" *Regional Studies* **35** 809 – 822
Richardson H W, 1985, "Input – output and economic base multipliers: looking backward and forward" *Journal of Regional Science* **25** 607 – 662
Sonis M, Hewings G J D, 1992, "Coefficient change in input – output models: theory and applications" *Economic Systems Research* **4** 143 – 157
Spijker J, 1985, "Een combinatie van de directe en indirecte methoden" [A combination of direct and indirect methods] in *Constructie en actualisering van regionale en interregionale input – output tabellen* Eds J Oosterhaven, B B A Drewes (Economisch-Technologische Instituten, Arnhem) pp 171 – 178
Stevens B H, Treyz G I, Lahr M L, 1989, "On the comparative accuracy of RPC estimation techniques", in *Frontiers of Input – Output Analysis* Eds R E Miller, K R Polenske, A Z Rose (Oxford University Press, Oxford) pp 245 – 257
Tiebout C M, 1967, "Input – output and the firm: a technique for using national and regional tables" *Review of Economics and Statistics* **49** 260 – 262
van Dijk J, Oosterhaven J, 1986, "Regional impacts of migrants' expenditures: an input – output/vacancy-chain approach", in *London Papers in Regional Science 15. Integrated Analysis of Regional Systems* Eds P W J Batey, M Madden (Pion, London) pp 122 – 147
West G R, 1990, "Regional trade estimation: a hybrid approach" *International Regional Science Review* **13** 103 – 118