Analytical and empirical comparison of policy-relevant key sector measures

Temurshoev, U.; Oosterhaven, J.

GGDC RESEARCH MEMORANDUM *132*

Analytical and Empirical Comparison of Policy-Relevant Key Sector Measures

Umed Temurshoev and Jan Oosterhaven

April 2013

university of groningen

groningen growth and development centre

Analytical and Empirical Comparison of Policy-Relevant Key Sector Measures*

Umed Temurshoev† Jan Oosterhaven‡

April 11, 2013

Abstract

We consider ten widely used key sector measures (linkages) and identify groups of the most similar indicators on both analytical and empirical grounds. We derive new closed-form formulas for the generalized complete and incomplete hypothetical extraction linkages and add the up till now undefined forward counterpart of the net backward linkage. The analytical relations and some stylized facts enable us to formulate hypotheses about the direction and strength of the relationships between various linkages. To study policy-relevant measures, our empirical tests are based on sectoral income (GDP) linkages, CO_2-emission linkages and employment linkages for 34 industries and 33 countries. They show that the information on the ten key sector measures may be summarized by four measures.

Keywords: Backward linkages, forward linkages, hypothetical extraction, net linkages

JEL Classification Codes: C67, D57, R11

*The authors thank Randy Jackson and Geoffrey Hewings for useful comments. The views expressed in this paper are those of the authors and should not be attributed to the European Commission or its services.

†European Commission, Joint Research Centre, Institute for Prospective Technological Studies, s/n 41092 Seville, Spain. E-mail: umed.temurshoev@ec.europa.eu (Corresponding author)

‡Faculty of Economics and Business, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands. E-mail: j.oosterhaven@rug.nl

1 Introduction

In the fields of regional economics and development economics different measures have been proposed and extensively tested to identify key sectors – mostly defined as sectors with a high potential of spreading growth impulses throughout the economy (see Hirschman, 1958, for a first account, and Miller and Blair, 2009, for a recent overview). The proliferation of measures partly reflects methodological improvements, such as the replacement of the *direct* backward linkages (Chenery and Watanabe, 1958) with the *total* backward linkages measured by the column sums of the Leontief-inverse (Rasmussen, 1956), or the replacement of the row sums of the Leontief-inverse (Rasmussen, 1956) with the row sums of the Ghosh-inverse (Beyers, 1976; Jones, 1976) in the case of total forward linkages.[1]

Partly, however, the proliferation is due to the different labeling of the same phenomena in independently written and/or seemingly unrelated studies. Thus, we have the output-to-output multiplier (Miller and Blair, 1985, 2009,p.328), which is equivalent to the total flow multiplier (Szyrmer, 1984, 1992), which is equivalent to the earlier hypothetical extraction (HE) of whole sectors (Paelinck et al., 1965; Strassert, 1968; Schultz, 1977). The last point was first indicated by Szyrmer (1992) and recently proved by Gallego and Lenzen (2005) and Temurshoev (2010a).[2] However, HE offers more flexibility as it allows to extract any subset of transactions instead of deleting only full rows *and* columns (see Miller and Lahr, 2001).[3]

[1] This is an improvement because the row sums of the Leontief-inverse define the backward impact of an economically meaningless unit vector of final demand. Recently, Antràs et al. (2012) proved that the total forward linkage quantifies an industry's average distance from its final output users, thus the measure indicates an industry's relative position in the output supply chain. Similarly, Miller and Temurshoev (2012) show that the total backward linkage is equivalent to the industry average distance from its primary inputs suppliers in the economy-wide input demand chain.

[2] Note that the prime purpose of Temurshoev (2010a) was formulating and deriving an analytical solution (referred to as a group factor worth, yet another label) to the problem of finding a key group of sectors with the HE method. It is a generalization of the HE key sector problem, but the author shows that the two problems are generally not equivalent.

[3] This flexibility is important, for example, in examining the effect of complex production pro-

Earlier studies used key sector measures predominantly in terms of gross output. However, to be really relevant to policy formulation, key sectors should be defined by means of measures that reflect main policy goals, such as income generation, job creation or reduction of CO_2-emissions, instead of being defined by total output (see Oosterhaven, 1981, ch. 5, for an early application of forward and backward employment linkages, and Lenzen, 2003, for a general discussion).

It is remarkable to note that the majority of all key sector measures try to capture the same basic concept: namely the dependence of the rest of the economy on the sector at hand in terms of the indicator chosen (i.e., output, employment, income, CO_2, etc.). The only exception seem to the *net* backward linkage interpretation (Oosterhaven, 2004; Dietzenbacher, 2005) of the net multiplier concept of Oosterhaven and Stelder (2002). As opposed to the standard key sector measures, the net linkage concept tries to capture the two-sided nature of sectoral dependence by taking the ratio of the dependence of the Rest of the Economy (RoE) on the sector at hand to the dependence of that sector on the RoE.

In view of the proliferation of key sector measures, this article tries to establish, both analytically and empirically, how (dis)similar these measures are. That is, the purpose of this paper is finding groups of similar key sector measures on the basis of analytical and empirical comparisons. Hence, our results are expected to be useful for the input-output practitioners and analysts in their choice of the smallest possible number of key sector measures.

We compare the ten most prominent key sector measures, which include direct, total, complete HE, incomplete HE and net backward linkages, as well as their

cesses on the output decline in transition economies. Blanchard and Kremer (1997), e.g., argue that the breakdown of complex chains of production caused by transition in former Soviet Union and Central European countries is one of the main factors explaining the rapid decline in their GDP from 1989 to 1994. Their empirical evidence suggests that output has fallen most for goods with the most complex production process.

forward equivalents. To study policy-relevant measures, we do not compare total output linkages, but income (GDP) linkages, CO_2-emission linkages and employment linkages. The average aggregate results are given for 33 countries, while the country-specific comparisons include twenty separate economies for 2005, representing a continuum of poor and rich, and big and small countries.

Section 2 summarizes and further discusses nine known output-based linkages, and adds the up till now undefined forward counterpart of the net backward linkage. Section 3 summarizes the known and unknown analytical relations between the generalized versions of these ten output-based linkages, and adds new closed-form formulas for the complete and incomplete HE backward and forward linkages. Section 4 empirically tests the (dis)similarity of the ten linkages for three factors (income, CO_2 and jobs) by means of the averages and the standard deviations of the 9x10 correlation coefficients for 33 countries, and further tests the mentioned (dis)similarities within 20 economies separately by means of hierarchical cluster analysis. Section 5 concludes that there are at most four mutually more of less independent groups of linkages.

2 Overview of output-based linkage measures

For completeness' sake and to compare the more complex measures with the most simple ones, we include the direct backward and forward output-based linkages by industry (Chenery and Watanabe, 1958). These measures are based on the *input coefficients* a_{ij}, indicating the use of domestic intermediate outputs of industry i per unit of output of buying industry j, and the *output coefficients* b_{ij}, indicating the domestic intermediate sales to industry per unit of output of selling industry i. With these coefficients, *direct backward* and *direct forward linkages* are defined as, respectively:

$$b_i^d = \sum_k a_{ki} \quad \text{and} \quad f_i^d = \sum_k b_{ik}, \quad \forall i, \tag{1}$$

where superscript d stands for the *direct* linkages of industry i.

The by now standard measures of backward and forward linkages add the higher order relations between industries i and j to (1). They are derived from, respectively, the Leontief (1936, 1941) demand-driven input-output (IO) model and the Ghosh (1958) supply-driven IO model.[4] The solution of the well-known Leontief model, that is $\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{y}$, delivers a row vector with *total output multipliers* $(\mathbf{b}^t)' = \boldsymbol{\imath}'\mathbf{L}$, where $\mathbf{x}$ is a column with total output by industry, $\mathbf{I}$ the identity matrix, $\mathbf{A}$ the matrix with the above defined input coefficients, $\mathbf{y}$ a column with final demand, $\boldsymbol{\imath}$ a summation vector of ones, and $\mathbf{L}$ the Leontief-inverse $(\mathbf{I} - \mathbf{A})^{-1}$. The solution of the by now almost equally well-known Ghosh model, that is $\mathbf{x}' = \mathbf{v}'(\mathbf{I} - \mathbf{B})^{-1}$, delivers a column with *total input multipliers* $\mathbf{f}^t = \mathbf{G}\boldsymbol{\imath}$,[5] where additionally $\mathbf{x}'$ is a row with total input by industry, $\mathbf{v}'$ a row with total primary inputs (including imports) by industry, $\mathbf{B}$ the matrix with the above defined output coefficients, and $\mathbf{G}$ the Ghosh-inverse $(\mathbf{I}-\mathbf{B})^{-1}$. These total output and total input multipliers define the *total backward* and *total forward linkages* as, respectively:

$$b_i^t = \sum_k l_{ki} \quad \text{and} \quad f_i^t = \sum_k g_{ik}, \quad \forall i, \tag{2}$$

where superscript t stands for the *total* linkages of industry i.

[4]Following the usual notation, matrices are indicated by bold capitals, vectors by bold lowercases and scalars by italic lowercases. Vectors are columns by definition; transposition is indicated by a prime; and $\hat{\mathbf{x}}$ denotes the diagonal matrix with $\mathbf{x}$ on the main diagonal.

[5]Note that the term total input multiplier implies a causal interpretation as it refers to the Ghosh model from which these multipliers are derived. The term forward linkage, however, does not necessarily imply a causal interpretation. In as far as the policy user insists on giving forward linkages a causal interpretation, the Ghosh model should be interpreted as a price model according to Dietzenbacher (1997), and not as a quantity model, since Oosterhaven (1988, 1989) shows that the supply-driven quantity model involves implausible assumptions and consequences.

Hypothetical extraction (HE) methods define a series of less known key sector measures. The central idea of the classical HE method is that the hypothetical elimination of a complete industry allows one to estimate its contribution to the economy-wide total output. After nullifying the i-th row and column of the input coefficient matrix, denoted by $\mathbf{A}^{-i}$, and nullifying the i-th element of the final demand vector, denoted $\mathbf{y}^{-i}$, the vector of total output after extracting sector i is given by:

$$\mathbf{x}_l^{-i} = \mathbf{L}^{-i}\mathbf{y}^{-i} \quad \text{with} \quad \mathbf{L}^{-i} = (\mathbf{I} - \mathbf{A}^{-i})^{-1}, \tag{3}$$

where subscript l in (3) refers to the fact that the HE is implemented with the Leontief model. The difference between the system-wide total output before and after the extraction, $\boldsymbol{\imath}'\mathbf{x} - \boldsymbol{\imath}'\mathbf{x}_l^{-i}$, indicates the total *backward impact* of the *complete extraction* of industry i (see Miller and Blair, 2009,ch. 12).

The forward impact of the elimination of industry i may be calculated in the same way with the Ghosh model. Note that in a causal sense both impacts cannot occur at the same time, as both models cannot be true at the same time. This means that when used together, the linkage measures derived from these two models should not be given a causal interpretation. Define $\mathbf{B}^{-i}$ as the output coefficient matrix with a nullified i-th row and column, and $(\mathbf{v}^{-i})'$ as the total primary inputs row vector with a nullified i-th element. Then, the value of the economy-wide total input after the elimination of sector i equals:

$$(\mathbf{x}_g^{-i})' = (\mathbf{v}^{-i})'\mathbf{G}^{-i} \quad \text{with} \quad \mathbf{G}^{-i} = (\mathbf{I} - \mathbf{B}^{-i})^{-1}, \tag{4}$$

where subscript g indicates a HE with the Ghosh model. Consequently, the *forward impact* of the *complete extraction* of industry i on the value of total inputs then equals $\mathbf{x}'\boldsymbol{\imath} - (\mathbf{x}_g^{-i})'\boldsymbol{\imath}$ (see Miller and Blair, 2009,ch. 12).

However, comparing *absolute HE impacts* is not very useful. They simply tell us that extracting big/small industries/regions/countries tends to have big/small impacts on the aggregate economy. Moreover, since we want to analytically and empirically compare the outcomes of different key sector measures, comparing the absolute HE impacts with (1) and (2) is senseless, as absolute HE impacts have a unit of measurement (of say, millions of euros or kilotons of CO_2), whereas the other linkages are dimensionless indicators. Hence, we will only consider *normalized* HE linkages that are expressed per unit of input or output. Thus, industry i's *backward* and *forward linkages* due to its *complete extraction* equal, respectively, its total backward impact, $\boldsymbol{\imath}'\mathbf{x} - \boldsymbol{\imath}'\mathbf{x}_l^{-i}$, and its total forward impact, $\mathbf{x}'\boldsymbol{\imath} - (\mathbf{x}_g^{-i})'\boldsymbol{\imath}$, divided by its own total output and input:[6]

$$b_i^c = \frac{\boldsymbol{\imath}'\mathbf{x} - \boldsymbol{\imath}'\mathbf{x}_l^{-i}}{x_i} \quad \text{and} \quad f_i^c = \frac{\mathbf{x}'\boldsymbol{\imath} - (\mathbf{x}_g^{-i})'\boldsymbol{\imath}}{x_i}, \quad \forall i, \tag{5}$$

where superscript c stands for the *complete* extraction of industry i.

As said before, the advantage of HE linkages is their flexibility despite their somewhat cumbersome calculation by means of (5). Temurshoev (2010a), for example, considers the difference between the sum of the backward linkages due to the HE of separate industries and the corresponding result due to the HE of a group of the same industries for such factors as water, CO_2 emissions, profits, and net wages. The author then interprets the obtained values of such differences as the degree of *redundancy* of sectors in terms of their linkage patterns, factor generation and final demand structures. Of the many other possibilities discussed in Miller and Lahr (2001), we only consider the application by Dietzenbacher and van der Linden

[6] Other normalizations are possible, such as with regard to aggregate total output (see Table 12.5 in Miller and Blair, 2009). However, for our empirical test normalization with respect to sectoral outputs, x_i, is most appropriate, since it avoids artificially high correlation between different IO linkages.

(1997), as their objective comes closest to the idea of finding the industries that have the largest potential to stimulate the economy as a whole. Their concept is similar to the complete extraction of a single industry, but they *only* nullify a column of the input matrix $\mathbf{A}$ (resp. a row of the output matrix $\mathbf{B}$) to quantify the backward (resp. forward) impact of the extracted industries.

More formally, let $\mathbf{A}_c^{-i}$ indicate the input coefficient matrix with zeros in column i, and $\mathbf{B}_r^{-i}$ the output coefficient matrix with zeros in row i.[7] Then, using the Leontief model to estimate the backward impacts and the Ghosh model to estimate the forward impacts, the corresponding *absolute impacts* of the extraction of industry i equal $\boldsymbol{\imath}'\mathbf{x} - \boldsymbol{\imath}'\mathbf{x}_c^{-i}$ and $\mathbf{x}'\boldsymbol{\imath} - (\mathbf{x}_r^{-i})'\boldsymbol{\imath}$, respectively, where $\mathbf{x}_c^{-i} = (\mathbf{I} - \mathbf{A}_c^{-i})^{-1}\mathbf{y}$ and $(\mathbf{x}_r^{-i})' = \mathbf{v}'(\mathbf{I} - \mathbf{B}_r^{-i})^{-1}$. The *normalized backward* and *forward linkages* of the *incomplete extraction* of industry i are then defined as:

$$b_i^i = \frac{\boldsymbol{\imath}'\mathbf{x} - \boldsymbol{\imath}'\mathbf{x}_c^{-i}}{x_i} \quad \text{and} \quad f_i^i = \frac{\mathbf{x}'\boldsymbol{\imath} - (\mathbf{x}_r^{-i})'\boldsymbol{\imath}}{x_i}, \quad \forall i, \tag{6}$$

where superscript i stands for the *incomplete* extraction of industry i.

More recently, Oosterhaven and Stelder (2002) proposed the *net multiplier* concept. When multiplied with the sectoral total outputs of all industries, it exactly reproduces the economy-wide total output, instead of a value that would be twice as high if standard Leontief multipliers of around 2.00 would have been used, which is what is done in many industry lobby reports. After a fierce debate (De Mesnard, 2002, 2007; Dietzenbacher, 2005; Oosterhaven, 2004, 2007) it was agreed that the net multiplier should be interpreted and labeled as a *net backward linkage.* It is

[7]Note that while $\mathbf{A}^{-i}$ means that both the i-th row and column of the original input matrix $\mathbf{A}$ are nullified, $\mathbf{A}_c^{-i}$ indicates that only its column i is set to zero. Analogously, when only the i-th row of $\mathbf{B}$ is nullified, the new output coefficient matrix is denoted by $\mathbf{B}_r^{-i}$.

defined as:

$$b_i^n = b_i^t \frac{y_i}{x_i}, \tag{7}$$

where the superscript n stands for the *net* linkage.

Dietzenbacher (2005) showed that (7) equals the i-th column sum of $\mathbf{L}\hat{\mathbf{y}}$ divided by the i-th row sum of $\mathbf{L}\hat{\mathbf{y}}$, i.e., it equals the output generated in all industries due to the final demand of industry i normalized by the output generated in industry i due to the final demand of all industries. Thus, being a key sector with $b_i^n > 1$ implies that the Rest of the Economy (RoE, including itself) is more dependent on industry i than industry i is dependent on the RoE (including itself), whereas $b_i^n < 1$ indicates the reverse situation. This emphasizes that the net backward linkage (7) measures the two-sided nature of sectoral dependency.[8] Note that compared to the other backward linkages, the *net* backward linkage of industry i can be negative in the rare case when the final demand of industry i is negative. However, also in that case this interpretation is still valid.

To get a complete set of forward and backward linkages, we here define and add to the literature the corresponding *net forward linkage* as:

$$f_i^n = f_i^t \frac{v_i}{x_i}, \tag{8}$$

It is easily shown that (8) equals the i-th row sum of $\hat{\mathbf{v}}\mathbf{G}$ normalized by the i-th column sum of $\hat{\mathbf{v}}\mathbf{G}$, i.e., it equals the gross inputs of all industries that utilize

[8]Note that one might think of all the *normalized* HE indicators as measuring such two-sided nature of dependencies as well. For example, while the numerator of b_i^c in (5) represents the impact of industry i on the entire economy, the denominator could be, similar to (7), interpreted as the impact of the entire economy on industry i. However, such interpretation is incorrect because the interpretation of the denominator is not consistent with the HE philosophy. In order to capture such two-sided nature of dependency and be comparable with the net linkage philosophy, the HE measures should divide the economy-wide impact of industry i's disappearance from the system by the impact of disappearing of all other sectors on industry i. But the last HE case makes little sense from practical point of view, though it can be quantified.

the primary input of industry i normalized by the gross input of industry i that absorbes/embodies the primary inputs of all industries. This implies that a key sector with $f_i^n > 1$ is an industry that is less dependent on the primary inputs of the RoE than the RoE is dependent on its primary inputs. Note that the value of the net forward linkage of industry i becomes negative in a rare case of a negative total primary inputs of industry i, due to large subsidies or economic losses. However, also in that case the above interpretation is still valid.

3 Analytical relationships between generalized linkages

Next, we consider the analytical relationships between the more policy-relevant, generalized versions of the above defined output-based linkages. The generalization may be based on any input or output factor (such as value added, employment, CO_2 emissions or energy-use) that may be linked to total output by industry (cf. Lenzen, 2003). The core equation of the generalization is $\pi = \boldsymbol{\pi}'\mathbf{x}$, in which π indicates the economy-wide use/emission of the factor at hand, while $\boldsymbol{\pi}'$ represents a row vector with *direct factor coefficients* indicating the sectoral factor use/emission per unit of sectoral total output. The equivalent output-based analogs of the coming results can simply be obtained by setting $\boldsymbol{\pi} = \boldsymbol{\imath}$ in the generalized linkage equations, as $x = \boldsymbol{\imath}'\mathbf{x}$ defines the economy-wide total output.

Adding factor coefficients to (1), and dividing by the direct factor coefficient π_i to obtain a dimensionless key sector measure, delivers the *normalized direct* backward and forward *factor linkages* of industry i:

$$b_i^{\pi,d} = \frac{\sum_k \pi_k a_{ki}}{\pi_i} \quad \text{and} \quad f_i^{\pi,d} = \frac{\sum_k \pi_k b_{ik}}{\pi_i}, \quad \forall i. \tag{9}$$

When π stands for the number of jobs, (9) measures the number of jobs needed in the industries supplying industry i per direct job in industry i in case of the direct backward linkages, whereas it measures the number of jobs in all purchasing industries tied to one direct job in industry i in case of the direct forward linkages.

The total factor multipliers of the Leontief model equal $\boldsymbol{\pi}'(\mathbf{I}-\mathbf{A})^{-1} = \boldsymbol{\pi}'\mathbf{L}$, while the total factor multipliers of the Ghosh model equal $(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\pi} = \mathbf{G}\boldsymbol{\pi}$. Dividing these total factor multipliers by the direct factor coefficient of industry i gives the *normalized total* backward and forward *factor linkages* of industry i as (cf. Oosterhaven, 1981):

$$b_i^{\pi,t} = \frac{\sum_k \pi_k l_{ki}}{\pi_i} \quad \text{and} \quad f_i^{\pi,t} = \frac{\sum_k \pi_k g_{ik}}{\pi_i}, \quad \forall i. \tag{10}$$

When π stands for the number of jobs and when (10) is given a causal interpretation, the numerator of (10) (i.e., non-normalized total backward factor linkage) indicates the economy-wide number of jobs generated per unit of final demand for products of industry i. Hence, its normalized counterpart (10) gives the economy-wide number of jobs per direct job in industry i. The non-normalized total forward factor linkage of industry i is best interpreted more loosely as the economy-wide number of jobs that is related to a unit of primary inputs of industry i. The normalization in (10) again gives this number per direct job in industry i.

The relation between the normalized total backward and forward factor linkages (in matrix form written, respectively, as $(\mathbf{b}_\pi^t)' = \boldsymbol{\pi}'\mathbf{L}\hat{\boldsymbol{\pi}}^{-1}$ and $\mathbf{f}_\pi^t = \hat{\boldsymbol{\pi}}^{-1}\mathbf{G}\boldsymbol{\pi}$) may be clarified by multiplying them with the direct factor use embodied in, respectively, actual final demand and actual primary input (cf. Oosterhaven, 1996):

$$(\mathbf{b}_\pi^t)'(\hat{\boldsymbol{\pi}}\mathbf{y}) = \boldsymbol{\pi}'\mathbf{L}\hat{\boldsymbol{\pi}}^{-1}\hat{\boldsymbol{\pi}}\mathbf{y} = \boldsymbol{\pi}'\mathbf{x} = \mathbf{x}'\boldsymbol{\pi} = \mathbf{v}'\hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^{-1}\mathbf{G}\boldsymbol{\pi} = (\mathbf{v}'\hat{\boldsymbol{\pi}})\mathbf{f}_\pi^t. \tag{11}$$

This shows that the total factor requirement needed to satisfy final demand $\mathbf{y}$ is equal to total factor use accompanying sectoral primary input $\mathbf{v}'$.

Furthermore, it is important to note the partial definitional relation between the normalized direct and total linkages that follows from the Taylor-expansion of the Leontief-inverse and the Ghosh-inverse. Applied to the normalized backward and forward *factor* linkages, this leads to the following relations:

$$(\mathbf{b}_\pi^t)' = \boldsymbol{\imath}' + (\mathbf{b}_\pi^d)' + (\mathbf{b}_\pi^d)'\hat{\boldsymbol{\pi}}(\mathbf{L}-\mathbf{I})\hat{\boldsymbol{\pi}}^{-1} \quad \text{and} \quad \mathbf{f}_\pi^t = \boldsymbol{\imath} + \mathbf{f}_\pi^d + \hat{\boldsymbol{\pi}}^{-1}(\mathbf{G}-\mathbf{I})\hat{\boldsymbol{\pi}}\mathbf{f}_\pi^d. \quad (12)$$

Hence, we *may* expect a *positive* correlation between the direct and the total linkages (i.e., between $\mathbf{b}_\pi^d$ and $\mathbf{b}_\pi^t$, and between $\mathbf{f}_\pi^d$ and $\mathbf{f}_\pi^t$). However, without an empirical study it is not clear how strong these correlations will be, because the size of the higher order linkages, reflected by $\hat{\boldsymbol{\pi}}(\mathbf{L}-\mathbf{I})\hat{\boldsymbol{\pi}}^{-1}$ and $\hat{\boldsymbol{\pi}}^{-1}(\mathbf{G}-\mathbf{I})\hat{\boldsymbol{\pi}}$, may be significant and may be quite different across sectors.

The net backward linkage was originally already formulated as a net factor multiplier $\boldsymbol{\pi}'\mathbf{L}\hat{\mathbf{y}}\hat{\mathbf{x}}^{-1}\hat{\boldsymbol{\pi}}^{-1}$ (Oosterhaven and Stelder, 2002). Here we will label this concept the *net backward factor linkage*, and define the new *net forward factor linkage* accordingly:

$$b_i^{\pi,n} = \frac{\sum_k \pi_k l_{ki} y_i}{\pi_i x_i} = b_i^{\pi,t}\frac{y_i}{x_i} \quad \text{and} \quad f_i^{\pi,n} = \frac{\sum_k v_i g_{ik}\pi_k}{\pi_i x_i} = f_i^{\pi,t}\frac{v_i}{x_i}, \quad \forall i. \quad (13)$$

The net backward factor linkage thus equals the amount of factor π generated in all sectors due to the final demand for products of industry i divided by the amount of factor π generated in industry i due to the final demand of all industries. Consequently, being a *net backward key sector* with $b_i^{\pi,n} > 1$ implies that the RoE, in terms of factor usage, is more dependent on industry i than industry i is dependent on the RoE, while $b_i^{\pi,n} < 1$ indicates the reverse case. This ratio emphasizes the two-sided nature of sectoral dependency in terms of factor use/emissions. Analogously, being a *net forward key sector* with $f_i^{\pi,n} > 1$ implies that the economy-wide factor usage related to the primary inputs of industry i is larger than the amount of factor π in industry i that is related to the primary inputs of all sectors. Hence, $f_i^{\pi,n} > 1$ indicates that the RoE, in terms of factor usage, is more dependent on industry i than industry i is dependent on the RoE, while $f_i^{\pi,n} < 1$ indicates the reverse case. Note that with $\pi_i = \pi_k = 1$ the net factor linkage definitions in (13) boil down to their corresponding output-based net linkages given in (7) and (8).

Note that the column sums of the intermediate and primary *input* coefficients equal one, that is $b_i^d + v_i/x_i = 1$, and that the same holds for the row sums of the intermediate and final *output* coefficients, that is $f_i^d + y_i/x_i = 1$. This means that there is a definitional relationship between the net backward factor linkage and the total backward factor and direct forward output linkages, while an analogous relationship exists for the net forward factor linkage:

$$b_i^{\pi,n} = b_i^{\pi,t}(1 - f_i^d) \quad \text{and} \quad f_i^{\pi,n} = f_i^{\pi,t}(1 - b_i^d) \tag{14}$$

The ratios y_i/x_i and v_i/x_i, in real life, take values across industries usually within the intervals [0.35, 0.85] and [0.50, 0.70], respectively.[9] Hence, we expect, especially, the net and total backward factor linkages to convey rather different outcomes.

The analytical relationships between the above three non-HE linkages are relatively straightforward. The convoluted three step extraction procedure for the complete and the incomplete HE measures, however, makes their analytical comparison with other factor linkages impossible, unless a closed form is found for them. Temurshoev (2010a) gives the analytical expression for the reduction in factor usage

[9]These intervals, and the ones to follow, are the means of the 15th and 85th percentiles of the corresponding data for 33 countries used in the empirical part of this paper.

due to the complete extraction of sector i from the Leontief system, and shows that $\boldsymbol{\pi}'\mathbf{x}-\boldsymbol{\pi}'(\mathbf{I}-\mathbf{A}^{-i})^{-1}\mathbf{y}^{-i} = \sum_k \pi_k l_{ki} x_i / l_{ii}$ (see Szyrmer, 1984, and Gallego and Lenzen, 2005, who derive a similar relationship for the HE *output* linkage). Its normalization by the total amount of factor use/emissions by industry i, $\pi_i x_i$, gives the *complete HE backward factor linkage* of industry i, which may thus very simply be calculated as:[10]

$$b_i^{\pi,c} = \frac{\boldsymbol{\pi}'\mathbf{x} - \boldsymbol{\pi}'\mathbf{x}_l^{-i}}{\pi_i x_i} = \frac{b_i^{\pi,t}}{l_{ii}}. \tag{15}$$

Equation (15) shows that the complete HE factor linkage equals its total factor linkage, $b_i^{\pi,t}$, normalized by its self-dependency as indicated by l_{ii}. This makes sense as a sector with a large normalized backward factor multiplier contributes relatively much to the economy-wide use/emissions of factor π, but a sector that is largely dependent on itself will have less potential of spreading exogenous impulses throughout the economy.[11]

The forward factor *impact* due to the complete elimination of industry i is, of course, derived with the Ghosh model. In the Appendix it is shown to equal $\mathbf{x}'\boldsymbol{\pi} - (\mathbf{v}^{-i})'(\mathbf{I}-\mathbf{B}^{-i})^{-1}\boldsymbol{\pi} = \sum_k x_i g_{ik} \pi_k / g_{ii}$. The Appendix also shows that the diagonal elements of the Leontief and Ghosh inverses are equal, i.e., $g_{ii} = l_{ii}$. Hence, normalization with the total factor usage, $\pi_i x_i$, gives the *complete HE forward factor linkage* of industry i:

$$f_i^{\pi,c} = \frac{\mathbf{x}'\boldsymbol{\pi} - (\mathbf{x}_g^{-i})'\boldsymbol{\pi}}{\pi_i x_i} = \frac{f_i^{\pi,t}}{l_{ii}}. \tag{16}$$

From (15) and (16) it follows that the sectoral outcomes of the complete HE factor

[10]Note that such normalization assumes that factor use and gross output are related in the sense that $x_i > 0$ implies $\pi_i > 0$, and $x_i = 0$ implies $\pi_i = 0$. Hence, if $x_i = 0$ (or, equivalently, $\pi_i = 0$), we set $b_i^{\pi,c} = b_i^{\pi,t} = 0$, which is consistent with the definition of the absolute HE backward linkage. A similar note, if appropriate, is necessary for all other factor normalized linkages considered in this paper.

[11]An interesting property of the backward (and forward) complete HE factor linkages is that they are invariant to netting out of intra-industry transactions, i.e., the results will be the same in a *net* IO setting where the internal industry flows are set to zero. For details, see Temurshoev (2010b,ch. 6).

linkages should be generally quite similar to those of their corresponding total factor linkages, because the total input self-dependencies, l_{ii}'s, mostly vary between about 1.01 to 1.20, which implies a very narrow range for their inverse, $1/l_{ii} \in [0.83, 0.99]$.

Next consider the relation between the net factor linkages and the complete HE linkages. Combining (13), (15) and (16) gives:

$$\frac{b_i^{\pi,n}}{b_i^{\pi,c}} = \frac{y_i}{x_i} l_{ii} \quad \text{and} \quad \frac{f_i^{\pi,n}}{f_i^{\pi,c}} = \frac{v_i}{x_i} l_{ii}. \tag{17}$$

This implies that whenever the final demand-to-gross output (resp. primary inputs-to-gross input) ratios and the total input self-dependencies are identical across industries, the net bacward (resp. forward) factor linkage is equivalent to the complete HE backward (resp. forward) linkage. As noted above, y_i/x_i and v_i/x_i take values across industries usually within the intervals [0.35, 0.85] and [0.50, 0.70], respectively, while l_{ii} mostly varies between 1.01 to 1.20. Hence, the ratios $b_i^{\pi,n}/b_i^{\pi,c}$ and $f_i^{\pi,n}/f_i^{\pi,c}$ in the majority of cases vary, respectively, within [0.35, 1.02] and [0.51, 0.84]. This implies that the net backward (resp. forward) factor linkage is expected to be *positively* correlated with the complete HE backward (resp. forward) factor linkage, but the relation should be *stronger* between the corresponding forward linkages.

The interesting next question is whether there also exist closed-forms for the generalized linkages resulting from an *incomplete* extraction of industry i (cf. Equation (6)), and whether they are comparably simple, and the answer is yes. This means that we can do away with the convoluted three-step calculation procedure also in the case of an incomplete extraction of only a single column in the $\mathbf{A}$ matrix and a single row in the $\mathbf{B}$ matrix. In the Appendix we prove the following result for, respectively, the normalized *incomplete HE backward* and *forward factor linkages* of industry i:[12]

$$b_i^{\pi,i} = \frac{\boldsymbol{\pi}'\mathbf{x} - \boldsymbol{\pi}'\mathbf{x}_c^{-i}}{\pi_i x_i} = \frac{b_i^{\pi,t} - 1}{l_{ii}} \quad \text{and} \quad f_i^{\pi,i} = \frac{\mathbf{x}'\boldsymbol{\pi} - (\mathbf{x}_r^{-i})'\boldsymbol{\pi}}{\pi_i x_i} = \frac{f_i^{\pi,t} - 1}{l_{ii}}, \quad \forall i, \quad (18)$$

whenever $x_i \neq 0$ and set to $b_i^{\pi,i} = f_i^{\pi,i} = 0$ otherwise (i.e., for non-existent domestic industries which do not generate income, jobs, CO_2-emissions, etc.).

Note that taking the difference between the complete and the incomplete HE factor linkages, using equations (15), (16) and (18), shows that this difference is simply equal to the inverse of an industry's total input self-dependency:

$$b_i^{\pi,c} - b_i^{\pi,i} = 1/l_{ii} = 1/g_{ii} = f_i^{\pi,c} - f_i^{\pi,i}. \quad (19)$$

Given that in practice mostly $1/l_{ii} \in [0.83, 0.99]$ (see above), the differences of the HE linkages in (19) will be rather small. This implies that the both HE factor linkages (i.e., $b_i^{\pi,c}$ and $b_i^{\pi,i}$ on the one hand, and $f_i^{\pi,c}$ and $f_i^{\pi,i}$ on the other) are expected to be *strongly* and *positively* correlated.

Note that (19) also implies $b_i^{\pi,c} - f_i^{\pi,c} = b_i^{\pi,i} - f_i^{\pi,i}$, which means that if the complete HE backward and forward linkages are (not) closely related, so should be their incomplete HE counterparts. In addition, definitions (15) and (16) imply that:

$$\frac{b_i^{\pi,c}}{f_i^{\pi,c}} = \frac{b_i^{\pi,t}}{f_i^{\pi,t}}. \quad (20)$$

Hence, a close similarity (or large dissimilarity) of the normalized total backward and forward linkages immediately implies the same kind of relation for the complete HE backward and forward linkages.

[12]Szyrmer (1992) defines his total intermediate flow matrix as $\mathbf{T}^{int} = \mathbf{A}(\mathbf{I}-\mathbf{A})^{-1}\hat{\mathbf{L}}^{-1}\hat{\mathbf{x}}$ (p. 924, Eq. 7). Hence, Szymer's total intermediate coefficient matrix is $\mathbf{T}^{int}\hat{\mathbf{x}}^{-1} = \mathbf{A}(\mathbf{I}-\mathbf{A})^{-1}\hat{\mathbf{L}}^{-1} = (\mathbf{L}-\mathbf{I})\hat{\mathbf{L}}^{-1}$, whose i-th column sum gives the incomplete HE backward *output* linkage of industry i, the first expression in (18) with $\pi_i = 1$ (see also, Jeong, 1984).

Since the Leontief and Ghosh models are interrelated (see the Appendix and Equation (11)), and since all linkages are based on these two models, all backward linkages are related, in one way or another, to all forward factor linkages (see, e.g., Equation (14)). The Appendix shows that some of the relationships among the different factor linkages may be summarized as follows:[13]

$$b_i^{\pi,i} = b_i^{\pi,c}\left[1 - \frac{1}{b_i^{\pi,t}}\right] = b_i^{\pi,c}\left[1 - \frac{1-f_i^d}{b_i^{\pi,n}}\right] = b_i^{\pi,c}\left[1 - \frac{f_i^{\pi,t}(1-b_i^d)}{b_i^{\pi,t}f_i^{\pi,n}}\right], \tag{21}$$

$$f_i^{\pi,i} = f_i^{\pi,c}\left[1 - \frac{1}{f_i^{\pi,t}}\right] = f_i^{\pi,c}\left[1 - \frac{1-b_i^d}{f_i^{\pi,n}}\right] = f_i^{\pi,c}\left[1 - \frac{b_i^{\pi,t}(1-f_i^d)}{f_i^{\pi,t}b_i^{\pi,n}}\right]. \tag{22}$$

Table 1 summarizes the ten generalized IO linkages whose empirical (dis)similarity will be tested in the next Section.

Table 1: Summary of generalized dimensionless key sector measures

Abbr.	Name of measure	Linkage formula
bd	Direct backward factor linkage	$b_i^{\pi,d} = \sum_k \pi_k a_{ki}/\pi_i$
fd	Direct forward factor linkage	$f_i^{\pi,d} = \sum_k b_{ik}\pi_k/\pi_i$
bt	Total backward factor linkage	$b_i^{\pi} = \sum_k \pi_k l_{ki}/\pi_i$
ft	Total forward factor linkage	$f_i^{\pi} = \sum_k g_{ik}\pi_k/\pi_i$
bc	Complete HE backward factor linkage	$b_i^{\pi,c} = b_i^{\pi,t}/l_{ii}$
fc	Complete HE forward factor linkage	$f_i^{\pi,c} = f_i^{\pi,t}/l_{ii}$
bi	Incomplete HE backward factor linkage	$b_i^{\pi,i} = (b_i^{\pi,t}-1)/l_{ii}$
fi	Incomplete HE forward factor linkage	$f_i^{\pi,i} = (f_i^{\pi,t}-1)/l_{ii}$
bn	Net backward factor linkage	$b_i^{\pi,n} = b_i^{\pi,t}(y_i/x_i)$
fn	Net forward factor linkage	$f_i^{\pi,n} = f_i^{\pi,t}(v_i/x_i)$

Legend: π_i's indicate direct factor coefficients, a_{ij} and b_{ij} are domestic intermediate input and output coefficients, respectively, l_{ij} and g_{ij} are the entries from the Leontief and the Ghosh inverses, respectively, and x_i, y_i and v_i are, respectively, gross output, final demand and total primary inputs of sector i. Note that setting $\pi_k = \pi_i = 1$ defines the corresponding output-based linkages. HE stands for hypothetical extraction. Abbreviations (Abbr.) are used in the next section. Multiplication of bd, fd, bt and ft by π_i, and of bc, fc, bi and fi by $\pi_i x_i$ gives their corresponding *non-normalized* alternatives.

[13] One can also write the corresponding relations in terms of changes in linkage indicators. For example, taking total differential of the first expression in (21) in case of output yields $db_i^i = \frac{b_i^t-1}{b_i^t}db_i^c + \frac{b_i^c}{(b_i^t)^2}db_i^t$. Note that the usual comparative static analysis (e.g., fix $db_i^t = 0$ and see how db_i^i and db_i^c are interrelated) cannot be performed, as a change in one linkage measure implies simultaneous changes in all other linkages.

4 Empirical similarities

In our empirical test of the similarity of the ten factor linkages discussed in Section 3 we use the 2005 national IO tables constructed by the EU-funded World Input-Output Database (WIOD, www.wiod.org). These tables distinguish between domestic and imported intermediate deliveries, but only the first are used in computing the required input and output coefficient matrices. We include 33 WIOD-countries.[14] The national IO tables are valued at basic prices and expressed in current prices (millions of USD). There are 35 industries in the dataset, but we exclude the sector "Private households with employed persons" because for the overwhelming majority of countries it plays no role in the interindustry system. As the policy-relevant factors of interest, we focus on *income* (i.e., value added at basic prices plus taxes less subsidies on products, which sums up to GDP at market prices), CO_2 *emissions*, and the *number of persons engaged*. CO_2 emissions and employment data are obtained from the WIOD environmental and socio-economic accounts (for further details about WIOD, see Timmer, 2012).

The (dis)similarity analysis of the normalized income, CO_2 and employment linkages is based on Pearson's linear correlation coefficients. As there are 45 pairs of linkages from the 10 linkages studied, we are able to consider 45 correlations for each of 3 factors and 33 countries. Since this number is quite large, we first look at the simple arithmetic averages of these correlations. They are reported in Table 2 along with their standard deviations.[15] For the sake of readability, ratings icons are added

[14]These are: Australia, Austria, Belgium, Bulgaria, Brazil, Canada, China, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Indonesia, India, Ireland, Italy, Japan, Korea, Mexico, Netherlands, Poland, Portugal, Romania, Russia, Slovak Republic, Spain, Sweden, Turkey, Taiwan, United Kingdom and United States. We excluded the smallest EU27 countries.

[15]The results of the country-level *gross output-weighted* averages and standard deviations (which would adequately take into account the heterogeneity of countries in terms of their economic size) are very similar to those presented in Table 2, hence are not reported (but are available upon request from the authors).

to the table which represent the relative size of the correlation coefficients: totally white rating-charts indicate the lowest (negative) correlation coefficients, while more colored rating-bars are added to the icons for the relatively higher correlation values. Table 2 thus gives the average overall picture of the similarity of the normalized income, CO_2 and employment linkages of the 33 countries considered.

For all three considered factors, the following outcomes can be immediately observed from Table 2:

1. As expected, if only by their different names, the two groups of most strongly positively correlated linkages are the backward linkages {bd,bt,bc,bi,bn}, on the one hand, and the forward linkages {fd,ft,fc,fi,fn}, on the other hand.
2. The similarity between the total, the complete HE and the incomplete HE backward linkages, and similarity between the same forward linkages is particularly strong, as predicted by (15)-(16) and (19)-(20), while that between the direct and the total linkages is a little less strong, as predicted by the discussion of (12).
3. Net backward linkages are only weakly correlated with the other backward linkages, as predicted by the discussion of (13)-(14).
4. The same holds *mutatis mutandis* for the forward linkages, but to a lesser extent, as predicted in the discussion of (13)-(14).
5. Net linkages have the relatively highest positive correlations with the corresponding HE counterparts, where the correlation is stronger for {fc,fn} than for {bc,bn}, as predicted in the discussion of (17) (see also fn. 4).
6. Pairwise comparisons of the same backward and forward linkages for each factor reveals that the weakest link exists between the net backward and net forward linkages {bn,fn}, which was not discussed earlier.

Except for the first and the last points above, all other observations are as predicted.

Table 2: Average income, CO_2-emissions and employment linkages correlations

	bd	bt	bc	bi	bn	fd	ft	fc	fi
	Averages and standard deviations of income linkages correlations: 33 countries, 2005								
bt	0.99 (0.00)								
bc	0.93 (0.00)	0.93 (0.00)							
bi	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)						
bn	0.51 (0.06)	0.52 (0.06)	0.63 (0.04)	0.57 (0.05)					
fd	0.46 (0.04)	0.46 (0.03)	0.32 (0.05)	0.40 (0.04)	-0.39 (0.09)				
ft	0.45 (0.04)	0.45 (0.04)	0.30 (0.06)	0.39 (0.04)	-0.39 (0.09)	0.99 (0.00)			
fc	0.40 (0.05)	0.39 (0.05)	0.34 (0.06)	0.37 (0.05)	-0.37 (0.10)	0.95 (0.00)	0.95 (0.00)		
fi	0.43 (0.04)	0.42 (0.04)	0.32 (0.05)	0.38 (0.05)	-0.39 (0.10)	0.98 (0.00)	0.99 (0.00)	0.98 (0.00)	
fn	-0.07 (0.06)	-0.08 (0.06)	-0.16 (0.05)	-0.12 (0.05)	-0.62 (0.05)	0.75 (0.02)	0.77 (0.02)	0.80 (0.02)	0.79 (0.02)
	Averages and standard deviations of CO_2 emission linkages correlations: 33 countries, 2005								
bt	0.97 (0.00)								
bc	0.98 (0.00)	0.99 (0.00)							
bi	0.98 (0.00)	0.99 (0.00)	1.00 (0.00)						
bn	0.88 (0.01)	0.89 (0.01)	0.91 (0.00)	0.91 (0.00)					
fd	0.57 (0.06)	0.60 (0.05)	0.59 (0.05)	0.59 (0.05)	0.32 (0.07)				
ft	0.58 (0.05)	0.63 (0.05)	0.60 (0.05)	0.61 (0.05)	0.34 (0.07)	0.98 (0.00)			
fc	0.59 (0.05)	0.63 (0.05)	0.62 (0.05)	0.62 (0.05)	0.35 (0.07)	0.98 (0.00)	1.00 (0.00)		
fi	0.59 (0.05)	0.63 (0.05)	0.61 (0.05)	0.62 (0.05)	0.35 (0.07)	0.99 (0.00)	1.00 (0.00)	1.00 (0.00)	
fn	0.55 (0.06)	0.59 (0.06)	0.57 (0.06)	0.58 (0.06)	0.32 (0.08)	0.96 (0.00)	0.98 (0.00)	0.99 (0.00)	0.99 (0.00)
	Averages and standard deviations of employment linkages correlations: 33 countries, 2005								
bt	0.99 (0.00)								
bc	0.98 (0.00)	0.98 (0.00)							
bi	0.99 (0.00)	0.99 (0.00)	1.00 (0.00)						
bn	0.81 (0.01)	0.80 (0.02)	0.82 (0.01)	0.81 (0.01)					
fd	0.76 (0.03)	0.79 (0.03)	0.78 (0.03)	0.79 (0.03)	0.46 (0.08)				
ft	0.76 (0.03)	0.78 (0.03)	0.77 (0.03)	0.78 (0.03)	0.44 (0.08)	0.99 (0.00)			
fc	0.76 (0.04)	0.78 (0.03)	0.78 (0.03)	0.79 (0.03)	0.46 (0.08)	0.99 (0.00)	0.99 (0.00)		
fi	0.76 (0.04)	0.78 (0.03)	0.78 (0.03)	0.79 (0.03)	0.45 (0.08)	0.99 (0.00)	1.00 (0.00)	1.00 (0.00)	
fn	0.68 (0.05)	0.69 (0.05)	0.70 (0.05)	0.70 (0.05)	0.40 (0.09)	0.95 (0.01)	0.96 (0.00)	0.97 (0.00)	0.97 (0.00)

Note: The linkages abbreviations are given in Table 1. Totally uncolored (white) ratings icons indicate the lowest (negative) correlation coefficients, while more colored rating-bars are added to the icons for the relatively higher correlation values.

While the first observation should not come as a surprise (after all, backward and forward linkages are entirely different concepts), the last finding can also be explained. Using the definitions of the net linkages we have

$$\frac{b_i^{\pi,n}}{f_i^{\pi,n}} = \frac{b_i^{\pi,t}}{f_i^{\pi,t}} \cdot \frac{y_i}{v_i}. \tag{23}$$

From our data of 1122 (= 34 × 33) final demand-to-primary input ratios, y_i/v_i, we find that their 15-th and 85-th percentiles equal to 0.57 and 1.47, respectively. Hence, though the ratio $b_i^{\pi,t}/f_i^{\pi,t}$ in (23) is relatively stable, the ratio y_i/v_i varies a lot, and that explains why the net backward and net forward linkages are weaker and/or even negatively correlated compared to the other pairs of linkages.

Finally, note that in comparison to the income and CO_2 linkages, all the employment linkages are remarkably highly, positively correlated. This could be explained by the fact that the direct employment coefficients vary much less than the direct income and CO_2 coefficients. The differences between the 85-th and 15-th percentiles (i.e., the spreads) of the 1122 direct coefficients for income, CO_2-emissions and employment, are found to be equal in our data to 0.354, 0.552 and 0.037, respectively. We could have stopped with our comparison here since all correlations reported in Table 2 are highly statistically significant, except for only two (out of 45) income linkages.[16] However, it is also instructive to look at the individual country-level correlation matrices.

Reading all the country-level correlation matrices, however, is not practical. Hence, we visualize these relationships by means of a *hierarchical agglomerative cluster analysis* (HCA). On the basis of a distance matrix, which in our case is defined as one minus the Pearson's linear correlation between the linkages, the HCA

[16]These cases with relatively large standard deviations include the correlation results for {bd,fn} and {bt,fn} for income linkages.

identifies groups of linkages that are most similar. The HCA starts (in our case) from 10 clusters of size 1, and at each stage of the process finds the two "closest" (most alike) clusters and joins them together. Then, on the basis of newly computed distance matrices, this process continues until only one cluster of size 10 remains. We use the *average link* criteria for forming clusters, which computes the average distance between all pairs of objects in any two clusters r and s as follows

$$d(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} distance(x_{ri}, x_{si}), \tag{24}$$

where n_r is the number of objects in cluster r and x_{ri} is the i-th object in cluster r (for further details see e.g., Lattin et al., 2003,ch. 8). The derived hierarchical sequence of merging clusters is visually depicted by a tree diagram, also called a *dendrogram*. The resulting dendrograms for income, CO_2-emissions and employment linkages of twenty selected countries are given, respectively, in Figures 1, 2 and 3. These dendrograms consist of many inverse U-shaped lines connecting the factor linkages in a hierarchical tree, where the height of each U represents the distance between the two groups being connected.[17]

From Figure 1 we find a consistent outcome for income linkages for all twenty selected countries: if we choose a small enough distance (i.e., a sufficiently high correlation) in order to separate four groups of linkages, these four groups would always be {bd,bt,bc,bi}, {bn}, {fd,ft,fc,fi} and {fn}. It should be also mentioned that fn is generally closer to all other forward linkages than bn is to the other backward linkages,[18] as predicted in the discussion of (13)-(14). If we want to see only two clusters of linkages, then generally backward linkages get separated from the forward linkages. However, even with two groups we have five cases (for Australia,

[17]We use MATLAB cluster analysis toolbox in our implementation of the HCA technique.

[18]That is, in the majority of cases fn joins the group of forward linkages {fd,ft,fc,fi} sooner than bn joins {bd,bt,bc,bi}.

Figure 1: Dendrograms of income linkages

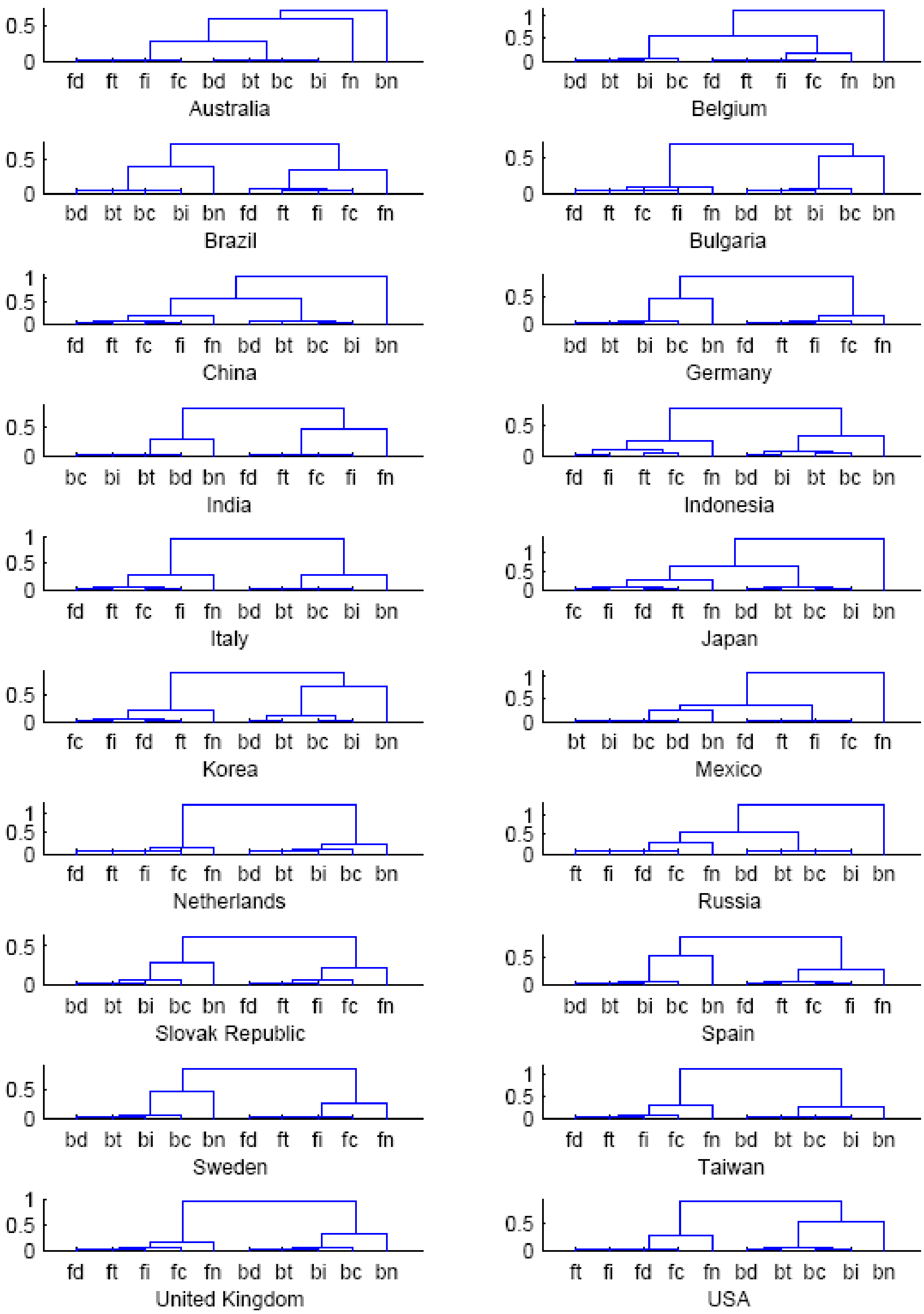


Figure 2: Dendrograms of CO_2 linkages

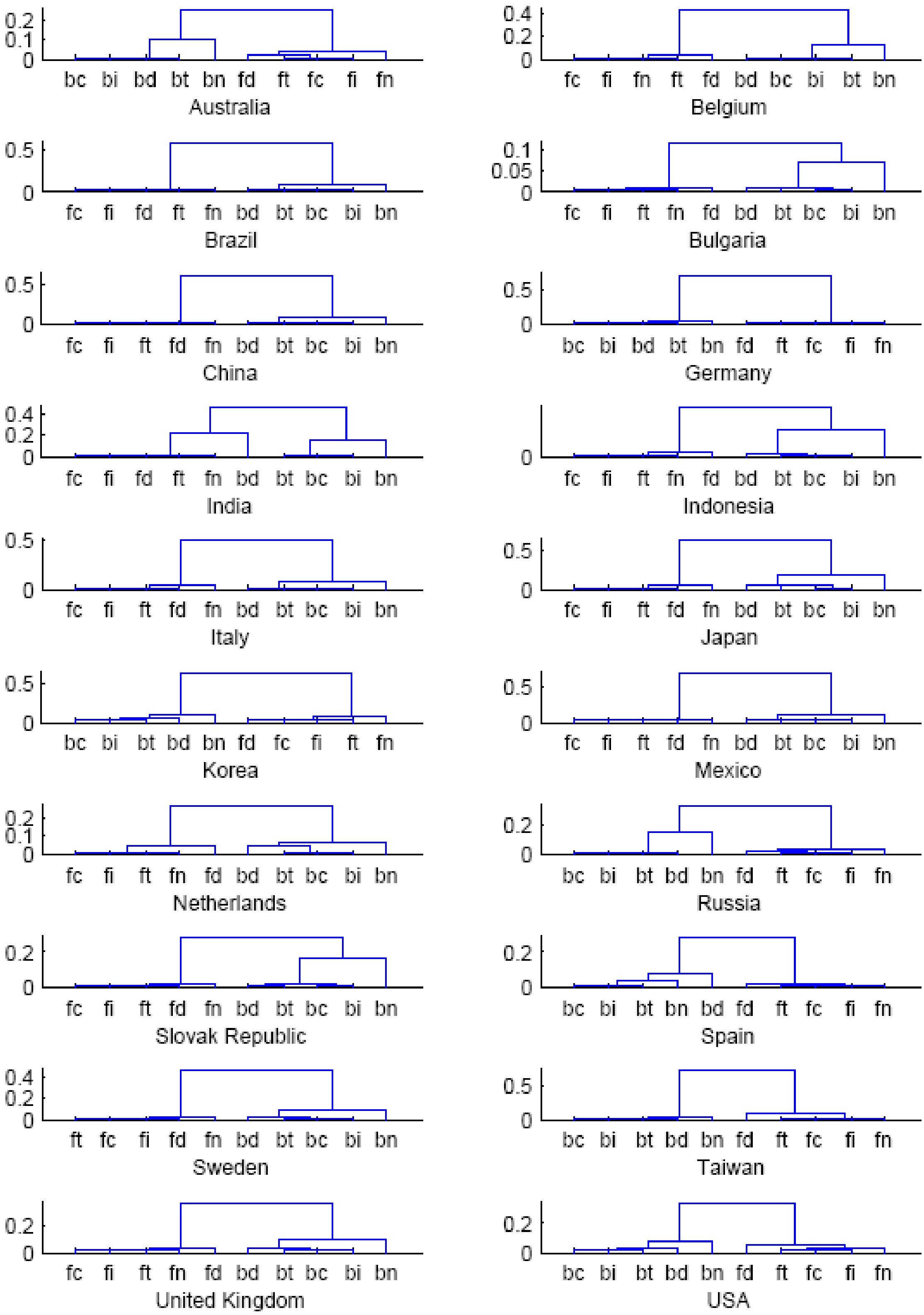
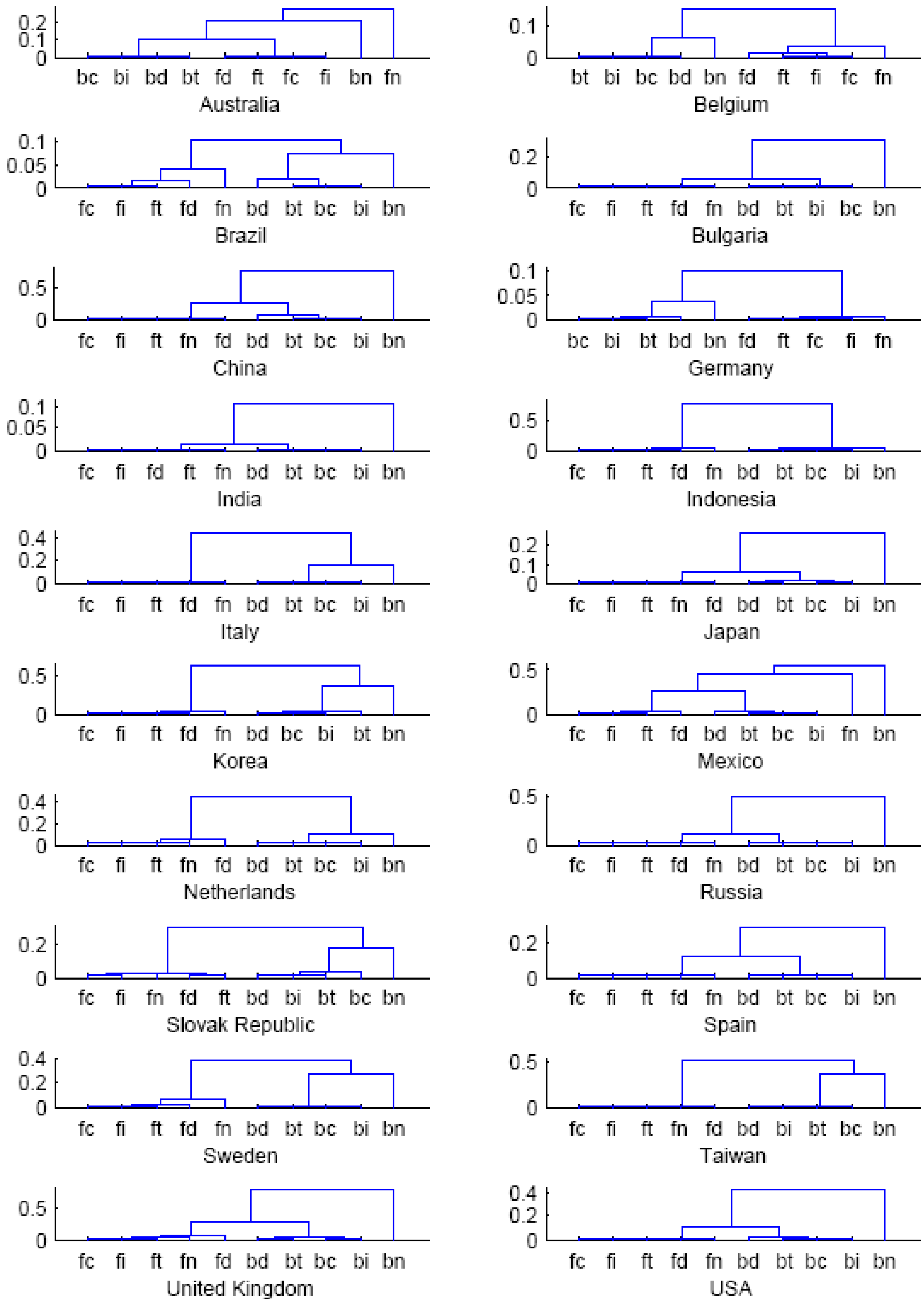


Figure 3: Dendrograms of employment linkages



Belgium, China, Japan and Russia) when bn is included solely in a separate group, and one case (for Mexico) when only fn makes up the other group.

For CO_2 linkages, if getting two groups of linkages is the goal, from Figure 2 we consistently find that backward linkages become separated from the forward linkages. If we want to further zoom in, then similar to the income linkages we find four separate groups {bd,bt,bc,bi}, {bn}, {fd,ft,fc,fi} and {fn} in seven cases (Australia, Italy, Japan, Korea, Russia, Slovak Republic and Sweden), while for seven other cases (Belgium, Bulgaria, Indonesia, Netherlands, Taiwan, United Kingdom and USA) the groups of forward linkages change to {ft,fc,fi,fn} and {fd}. However, it should be mentioned that all the forward linkages are very closely related, and thus the cluster separation in these last seven cases is probably irrelevant. But, also with CO_2-emission linkages, we clearly see again that the net backward linkage stands out from all other backward linkages.

As reported in Table 2, all employment linkages are strongly positively correlated. However, their dendrograms given in Figure 3 again zoom in further to find the groups of most closely related linkages.[19] For the majority of cases (namely, 12 countries) we could distinguish three clusters of similar linkages: {fd,ft,fc,fi,fn}, {bd,bt,bc,bi} and {bn}. In other cases, where one can separate four groups, fn becomes separated from the group of forward linkages, except for the Netherlands, where fn's place is taken by fd. All in all, here results are in accordance with those of the CO_2 linkages, where a further distinction between the forward linkages is irrelevant because they are all very strongly positively correlated.

[19]The original high correlations for CO_2 and employment linkages reported in Table 2 explain why the distance scale (which is equal to 1 minus correlation coefficient) of the corresponding dendrograms in Figures 2 and 3 are almost always less than 0.5.

5 Conclusion

The main purpose of this paper was finding on both analytical and empirical grounds which key sector measures are most similar and which are most dissimilar. We expect that the results of this study contribute to a better understanding of the relationships between a wide range of existing key sector indicators, some of them are, in fact, identical to each other and only have different labels in the literature, which may easily confuse input-output practitioners.

We compare the ten most prominent key sector measures, which include direct, total, complete hypothetical extraction (HE), incomplete HE and net backward linkages, as well as their forward equivalents. To study policy-relevant measures, we do not consider the traditional gross output-based linkages, but income (GDP) linkages, CO_2-emission linkages and employment linkages.

For comparison purposes we first summarize in detail the known and unknown analytical relations between the generalized key sector indicators, and add new closed-form formulas for the incomplete HE backward and forward linkages and also add the up till now undefined forward counterpart of the net backward linkage. Based on the derived analytical relations and on stylized facts, we are able to make predictions on the direction and strength of the relationships between different linkages. The test of these predictions in the empirical part of this paper are based on the 2005 data for 33 countries, representing a continuum of poor and rich, and big and small nations, while country-specific comparisons are made by means of a hierarchical cluster analysis for twenty separate economies.

The main finding from our analytical and empirical comparisons is that there are generally three clusters of similar factor linkages, which are:

1. the direct, the total, the complete HE and the incomplete HE *backward link-*

ages,

2. the net backward linkage, and
3. the direct, the total, the complete HE, the incomplete HE and the net *forward linkages*.

However, we also found that for income linkages, the net forward linkage becomes a key-sector indicator separate from the other forward linkages. Hence, the message for input-output practitioners is that they might simply use three linkages, i.e., choose one from each mentioned cluster, instead of the ten discussed key sector measures. In case of income linkages, it is also recommended to use two forward linkages, one from which being the net forward linkage. Of course, one may apply all the ten linkages, but we expect that the results of the measures belonging to the same cluster will be generally identical.

Finally, note that the basic similarities and dissimilarities summarized above are natural. All non-net linkages have the same basic philosophy, namely to measure the dependency of the rest of the aggregate economy on the industry/region/country at hand, either from a backward (i.e. purchasing) perspective or from the different forward (i.e. further processing) perspective. Hence, the outcomes should not be too different. In fact, both analytically and empirically we found them to be quite similar. The two net linkage concepts, on the other hand, have a basically different philosophy, namely to measure not the one-sided, but the two-sided nature of dependency between the industry/region/country at hand and the economy at large of which they are part. Hence, their outcomes should be different from those of the comparable non-net key sector indicators. From this perspective, it might be called a surprise that our new net forward linkage is less different from the other forward linkages than we expected it to be beforehand.

References

Antràs, P., D. Chor, T. Fally, and R. Hillberry: 2012, 'Measuring the upstreamness of production and trade flows'. *American Economic Review Papers and Proceedings* **102**, 412–416.

Beyers, W.: 1976, 'Empirical identification of key sectors: some further evidence'. *Environmental and Planning A* **8**, 231–236.

Blanchard, O. and M. Kremer: 1997, 'Disorganization'. *Quarterly Journal of Economics* **112**, 1091–1126.

Chenery, H. B. and T. Watanabe: 1958, 'Internatioinal comparisons of the structure of production'. *Econometrica* **26**, 487–521.

De Mesnard, L.: 2002, 'Note about the concept of net multipliers'. *Journal of Regional Science* **42**, 545–548.

De Mesnard, L.: 2007, 'A critical comment on Oosterhaven-Stelder net multipliers'. *Annals of Regional Science* **41**, 249–271.

Dietzenbacher, E.: 1997, 'In vindication of the Ghosh model: a reinterpretation as a price model'. *Journal of Regional Science* **37**, 629–651.

Dietzenbacher, E.: 2005, 'More on multipliers'. *Journal of Regional Science* **45**, 421–426.

Dietzenbacher, E. and J. van der Linden: 1997, 'Sectoral and spartial linkages in the EC production structure'. *Journal of Regional Science* **37**, 235–257.

Gallego, B. and M. Lenzen: 2005, 'A consistent input-output formulation of shared produce and consumer responsibility'. *Economic Systems Research* **17**, 365–391.

Ghosh, A.: 1958, 'Input-output approach to an allocative system'. *Economica* **25**, 58–64.

Henderson, H. V. and S. R. Searle: 1981, 'On deriving the inverse of a sum of matrices'. *SIAM Review* **23**, 53–60.

Hirschman, A.: 1958, *The Strategy of Economic Development*. New Haven: Yale University Press.

Jeong, K.: 1984, 'The relation between two different notions of direct and indirect input requirements'. *Journal of Macroeconomics* **6**, 473–476.

Jones, L. P.: 1976, 'The measurement of Hirschmanian linkages'. *Quarterly Journal of Economics* **90**, 323–333.

Lattin, J., J. D. Carrol, and P. E. Green: 2003, *Analyzing Multivariate Data*. London: Thomson Learning, Inc.

Lenzen, M.: 2003, 'Environmentally important paths, linkages and key sectors in the Australian economy'. *Structural Change and Economic Dynamics* **14**, 1–34.

Leontief, W. W.: 1936, 'Quantitative input-output relations in the economic system of the United States'. *Review of Economics and Statistics* **18**, 105–125.

Leontief, W. W.: 1941, *The Structure of American Economy, 1919-1929: An Empirical Application of Equilibrium Analysis.* Cambridge: Cambridge University Press.

Miller, R. E. and P. D. Blair: 1985, *Input-Output Analysis: Foundations and Extensions.* Engelwood Cliffs: Pretice-Hall, 1st edition.

Miller, R. E. and P. D. Blair: 2009, *Input-Output Analysis: Foundations and Extensions.* Cambridge: Cambridge University Press, 2nd edition.

Miller, R. E. and M. L. Lahr: 2001, 'A taxonomy of extractions'. In: M. L. Lahr and R. E. Miller (eds.): *Regional Science Perspectives in Economics: A Festschrift in Memory of Benjamin H. Stevens.* Amsterdam: Elsevier Science, pp. 407–441.

Miller, R. E. and U. Temurshoev: 2012, 'Upstreamness and downstreamness in world production'. *Mimeo.*

Oosterhaven, J.: 1981, *Interregional Input-Output Analysis and Dutch Regional Policy Problems.* Aldershot: Gower.

Oosterhaven, J.: 1988, 'On the plausibility of the supply-driven input-output model'. *Journal of Regional Science* **28**, 203–217.

Oosterhaven, J.: 1989, 'The supply-driven input-output model: a new interpretation but still implausible'. *Journal of Regional Science* **29**, 459–465.

Oosterhaven, J.: 1996, 'Leontief versus Ghoshian price and quantity models'. *Southern Economic Journal* **62**, 750–759.

Oosterhaven, J.: 2004, 'On the definition of key sectors and the stability of net versus gross multipliers'. Research Report 04C01, SOM Research School, University of Groningen.

Oosterhaven, J.: 2007, 'The net multiplier is a new key sector indicator: reply to De Mesnard's comment'. *Annals of Regional Science* **41**, 273–283.

Oosterhaven, J. and D. Stelder: 2002, 'Net multipliers avoid exaggerating impacts: with a bi-regional illustriation for the Dutch transportation sector'. *Journal of Regional Science* **42**, 533–543.

Paelinck, J., J. de Caevel, and D. J.: 1965, 'Analyse Quantitative de Certaines Phénomènes du Développment Régional Polarisé: Essai de Simulation Statique d'itérarires de Propogation'. In: *Problémes de Conversion Éconmique: Analyses Théoretiques et Études Appliquées.* Paris, pp. 341–387, M.-Th. Génin.

Rasmussen, P. N.: 1956, *Studies in Inter-Sectoral Relations.* Amsterdam: North-Holland.

Schultz, S.: 1977, 'Approaches to identifying key sectors empirically by means of input-output analysis'. *Journal of Development Studies* **14**, 77–96.

Strassert, G.: 1968, 'Zur bestimmung strategischer sektoren mit hilfe von von input-output modellen'. *Jahrbűcher fűr Nationalőkonomie und Statistik* **182**, 211–215.

Szyrmer, J. M.: 1984, *Total Flow in Input-Output Models.* PhD Thesis, School of Arts and Sciences, University of Pennsylvania, PA.

Szyrmer, J. M.: 1992, 'Input-output coefficeintsand multipliers from a total-flow perspective'. *Environmental and Planning A* **24**, 921–937.

Temurshoev, U.: 2010a, 'Identifying optimal sector groupings with the hypothetical extraction method'. *Journal of Regional Science* **50**, 872–890.

Temurshoev, U.: 2010b, *Interdependencies: Essays on Cross-Shareholdings, Social Networks, and Sectoral Linkages.* PhD Thesis, University of Groningen, Groningen: Ipskamp Drukkers B.V.

Timmer, M. P. (ed.): 2012, *The World Input-Output Database (WIOD): Contents, Sources and Methods*. Available at: `http://www.wiod.org/publications/source_docs/WIOD_sources.pdf`.

Appendix

Derivation of (16). The reduced outputs after extraction of sector i in the Ghosh framework are computed using $(\mathbf{x}^{-i})' = (\mathbf{v}^{-i})'\mathbf{G}^{-i}$, where $\mathbf{G}^{-i} = (\mathbf{I}-\mathbf{B}^{-i})^{-1}$. Note that because this is a complete extraction technique, both the i-th row and column of the output matrix $\mathbf{B}$ and the i-th element of the primary inputs vector $\mathbf{v}$ are nullified.

Lemma 1 (which is a particular case of Lemma 2) in Temurshoev (2010a) in terms of Ghosh inverse states that

$$\mathbf{G} - \mathbf{G}^{-i} = \frac{1}{g_{ii}}\mathbf{G}\mathbf{e}_i\mathbf{e}_i'\mathbf{G} - \mathbf{e}_i\mathbf{e}_i', \tag{A1}$$

where $\mathbf{e}_i$ is the i-th column of the identity matrix. Using (A1), Ghosh model and the fact that $\mathbf{v} - \mathbf{v}^{-i} = v_i\mathbf{e}_i$, we derive

$$\begin{aligned}
\mathbf{x}' - (\mathbf{x}^{-i})' &= \mathbf{v}'\mathbf{G} - (\mathbf{v}^{-i})'\mathbf{G}^{-i} \pm \mathbf{v}'\mathbf{G}^{-i} = \mathbf{v}'(\mathbf{G} - \mathbf{G}^{-i}) + (\mathbf{v}' - (\mathbf{v}^{-i})')\mathbf{G}^{-i} \\
&= \mathbf{v}'\left[\frac{1}{g_{ii}}\mathbf{G}\mathbf{e}_i\mathbf{e}_i'\mathbf{G} - \mathbf{e}_i\mathbf{e}_i'\right] + v_i\mathbf{e}_i'\left[\mathbf{G} - \frac{1}{g_{ii}}\mathbf{G}\mathbf{e}_i\mathbf{e}_i'\mathbf{G} + \mathbf{e}_i\mathbf{e}_i'\right] \\
&= \frac{x_i}{g_{ii}}\mathbf{e}_i'\mathbf{G} - v_i\mathbf{e}_i' + v_i\mathbf{e}_i'\mathbf{G} - \frac{v_i}{g_{ii}}g_{ii}\mathbf{e}_i'\mathbf{G} + v_i\mathbf{e}_i' \\
&= \frac{x_i}{g_{ii}}\mathbf{e}_i'\mathbf{G}.
\end{aligned} \tag{A2}$$

Hence, the complete HE forward factor linkage of sector i can be simply written as

$$f_i^{\pi,c} = \frac{(\mathbf{x}' - (\mathbf{x}^{-i})')\boldsymbol{\pi}}{\pi_i x_i} = \frac{x_i\mathbf{e}_i'\mathbf{G}\boldsymbol{\pi}}{\pi_i g_{ii} x_i} = \frac{f_i^{\pi,t}}{g_{ii}},$$

since the vector of total forward linkage is $\mathbf{f}_\pi^t = \hat{\boldsymbol{\pi}}^{-1}\mathbf{G}\boldsymbol{\pi}$. This proves (16) because the diagonal elements of the Leontief and Ghosh inverses are equal as follows from

$$\mathbf{L} = (\mathbf{I} - \mathbf{Z}\hat{\mathbf{x}}^{-1})^{-1} = (\mathbf{I} - \hat{\mathbf{x}}\mathbf{B}\hat{\mathbf{x}}^{-1})^{-1} = (\hat{\mathbf{x}}(\mathbf{I} - \mathbf{B})\hat{\mathbf{x}}^{-1})^{-1} = \hat{\mathbf{x}}\mathbf{G}\hat{\mathbf{x}}^{-1}, \tag{A3}$$

where $\mathbf{Z}$ is the intersectoral transaction matrix, that is, $l_{ii} = (x_i g_{ii})/x_i = g_{ii}$ for all i. $\square$

Derivation of (18). It is easy to confirm that $\mathbf{A}_c^{-i} = \mathbf{A}(\mathbf{I} - \mathbf{e}_i\mathbf{e}_i')$, where $\mathbf{e}_i$ is the i-th column of the identity matrix. In order to derive the reduced outputs, we make use of the following important identity that holds for any nonsingular matrix $\mathbf{X}$ and any vectors $\mathbf{u}$ and $\mathbf{v}$ (see Henderson and Searle, 1981,p. 53):

$$(\mathbf{X} + \mathbf{u}\mathbf{z}')^{-1} = \mathbf{X}^{-1} - \frac{1}{1 + \mathbf{z}'\mathbf{X}^{-1}\mathbf{u}}\mathbf{X}^{-1}\mathbf{u}\mathbf{z}'\mathbf{X}^{-1}. \tag{A4}$$

Choose $\mathbf{X} = \mathbf{I} - \mathbf{A}$, $\mathbf{u} = \mathbf{A}\mathbf{e}_i$ and $\mathbf{z} = \mathbf{e}_i$, hence we have

$$\begin{aligned}
\mathbf{x}_c^{-i} &= (\mathbf{I} - \mathbf{A}_c^{-i})^{-1}\mathbf{y} = (\mathbf{I} - \mathbf{A} + \mathbf{A}\mathbf{e}_i\mathbf{e}_i')^{-1}\mathbf{y} = \left[\mathbf{L} - \frac{1}{1 + \mathbf{e}_i'\mathbf{L}\mathbf{A}\mathbf{e}_i}\mathbf{L}\mathbf{A}\mathbf{e}_i\mathbf{e}_i'\mathbf{L}\right]\mathbf{y} \\
&= \mathbf{x} - \frac{1}{1 + \mathbf{e}_i'(\mathbf{L} - \mathbf{I})\mathbf{e}_i}(\mathbf{L} - \mathbf{I})\mathbf{e}_i\mathbf{e}_i'\mathbf{x} = \mathbf{x} - \frac{x_i}{l_{ii}}(\mathbf{L} - \mathbf{I})\mathbf{e}_i,
\end{aligned}$$

where we have used the fact that $\mathbf{L}\mathbf{A} = \mathbf{A} + \mathbf{A}^2 + \cdots = \mathbf{L} - \mathbf{I}$. This together with the definition of total backward factor linkage from (10) implies that the incomplete hypothetical extraction backward factor linkage of sector i is equal to

$$b_i^{\pi,i} = \frac{\boldsymbol{\pi}'\mathbf{x} - \boldsymbol{\pi}'\mathbf{x}_c^{-i}}{\pi_i x_i} = \frac{1}{\pi_i l_{ii}}\boldsymbol{\pi}'(\mathbf{L}-\mathbf{I})\mathbf{e}_i = \frac{b_i^{\pi,t}-1}{l_{ii}},$$

whenever $x_i = \pi_i \neq 0$, and $b_i^{\pi,i} = 0$ otherwise. This proves the first expression in (18).

Next employ the identity (A4) with $\mathbf{X} = \mathbf{I} - \mathbf{B}$, $\mathbf{u} = \mathbf{e}_i$ and $\mathbf{z}' = \mathbf{e}_i'\mathbf{B}$, and noting that $\mathbf{B}_r^{-i} = (\mathbf{I} - \mathbf{e}_i\mathbf{e}_i')\mathbf{B}$ we obtain

$$\begin{aligned}(\mathbf{x}_r^{-i})' &= \mathbf{v}'(\mathbf{I}-\mathbf{B}_r^{-i})^{-1} = \mathbf{v}'(\mathbf{I}-\mathbf{B}+\mathbf{e}_i\mathbf{e}_i'\mathbf{B})^{-1} = \mathbf{v}'\left[\mathbf{G} - \frac{1}{1+\mathbf{e}_i'\mathbf{B}\mathbf{G}\mathbf{e}_i}\mathbf{G}\mathbf{e}_i\mathbf{e}_i'\mathbf{B}\mathbf{G}\right] \\ &= \mathbf{x}' - \frac{1}{1+\mathbf{e}_i'(\mathbf{G}-\mathbf{I})\mathbf{e}_i}\mathbf{x}'\mathbf{e}_i\mathbf{e}_i'(\mathbf{G}-\mathbf{I}) = \mathbf{x}' - \frac{x_i}{l_{ii}}\mathbf{e}_i'(\mathbf{G}-\mathbf{I}),\end{aligned}$$

where we used the facts $\mathbf{B}\mathbf{G} = \mathbf{B} + \mathbf{B}^2 + \cdots = \mathbf{G} - \mathbf{I}$, and $g_{ii} = l_{ii}$ proved in (A3). Thus, for $x_i \neq 0$ the incomplete hypothetical extraction forward factor linkage of sector i is

$$f_i^{\pi,i} = \frac{\mathbf{x}'\boldsymbol{\pi} - (\mathbf{x}_r^{-i})'\boldsymbol{\pi}}{\pi_i x_i} = \frac{1}{\pi_i l_{ii}}\mathbf{e}_i'(\mathbf{G}-\mathbf{I})\boldsymbol{\pi} = \frac{f_i^{\pi,1}-1}{l_{ii}},$$

where we used the definition of the total forward factor linkage from (10). This proves the second expression in (18). Finally, note that whenever $x_i = \pi_i = 0$, we again set $f_i^{\pi,i} = 0$. □

Derivation of (21)-(22). Recall from (19) that $1/l_{ii} = b_i^{\pi,c} - b_i^{\pi,i}$. Plug this in the formula of $b_i^{\pi,i}$ in (18) and simple algebra gives $b_i^{\pi,i} = b_i^{\pi,c}(b_i^{\pi,t}-1)/b_i^{\pi,t}$, which is the first expression for $b_i^{\pi,i}$ in (21). From (14) it follows that $b_i^{\pi,t} = b_i^{\pi,n}/(1-f_i^d)$, which if plugged in the last expression gives the second formulation for $b_i^{\pi,i}$. Next, observe from (14) that the net backward and net forward linkages are related through

$$b_i^{\pi,n} = \frac{b_i^{\pi,t}(1-f_i^d)}{f_i^{\pi,t}(1-b_i^d)}f_i^{\pi,n}. \tag{A5}$$

Substituting (A5) in the second expression for $b_i^{\pi,i}$ gives its final formulation given in (21).

Similarly, from (19) we have $1/l_{ii} = f_i^{\pi,c} - f_i^{\pi,i}$. Plug this in the formula of $f_i^{\pi,i}$ in (18) and simple algebraic transformations give the first expression for $f_i^{\pi,i}$ in (22). From (14) it follows that $f_i^{\pi,t} = f_i^{\pi,n}/(1-b_i^d)$, which if plugged in the last expression gives the second formulation for $f_i^{\pi,i}$. Using (22) in the second expression for $f_i^{\pi,i}$ gives its final formulation in (22). Finally, note that (A5) is, in fact, another relation among the given factor linkages. □