Structural similarities between Input-Output tables: a comparison of OECD economies.

John Holt[a]*

[a]*Department of Mathematics, University of Auckland, Auckland, New Zealand*

john.holt@holtdatascience.co.nz

Structural similarities between Input-Output tables: a comparison of OECD economies.

We develop an approach to building similarity measures between input-output tables and apply it to the STAN collection of input-output tables maintained by the OECD, constructing a network graph whose vertices are input-output tables and where edges join vertices that are structurally similar. We call this graph the Economy Space, though it is not canonical.

Community detection techniques on Economy Space provide groups of economies that share structural similarities; interestingly from a path-dependence perspective, economies in the same group often have similar histories of political or social development.

The approach provides an analytical framework for the development of economic policy. We illustrate with an application to New Zealand.

JEL codes: O10,O11,O25, O57.

Keywords: input-output analysis, path dependence, structural decomposition analysis, community detection, economic development.

Introduction

A nation's government will generally seek to develop economic policy that promotes the nation's prosperity. It can be difficult to know when they have been successful, so policy-makers and governments will rely on macroeconomic indicators and measures to assist them in understanding whether the economy is becoming more prosperous, under some suitable definition of prosperous. Often these indicators are widely used, and so comparisons between economies can be made, and benchmarks can be constructed.

Governments and the public sector will adopt goals based on these measures and comparisons. For example, a country in the mid-to-lower ranks of the OECD rankings may desire to improve their measures so that they are in more prosperous company, though they may not necessarily aspire to be at the very top. Having identified the economies whose company they wish to join, a nation's policy makers might make comparative analyses to identify what should be done to make these changes: what legislation to adopt, what industries to support, how much funding should be apportioned to where?

Two countries may have a similar value for an economic indicator, but there might be very different reasons why the values are what they are. There may be very many mechanisms for an economy to achieve a certain value of economic indicator but it may be that none of these mechanisms are very close to the mechanism that operates in a given economy currently, and so it may not be feasible to achieve improvement in the near term.

This paper presents an approach to finding paths of economies connecting two economies. More precisely, if we approximate an economy by an Input-Output table, we develop an approach to finding paths of Input-Output tables joining any two Input-Output tables. The approach is to use a distance function to metrise the set of Input-Output tables, forming a weighted undirected graph, and then to prune this graph to remove edges between Input-Output tables that are far apart. We call the result *Economy Space* - it is not canonical, but each distance function provides a canonical one-parameter family of nested graphs, one of which is Economy Space. Figure 1 depicts the Economy Space for the Input-Output tables available from the OECD's STAN repository (OECD, 2016).

The distance between two Input-Output tables comes as a measure of difference between structural aspects of the two economies. When two Input-Output tables are near, there will be a set of sectors (possibly a different set for each economy) that represent the source and nature of the difference and this provides a starting point for analysis to inform policy development and economic planning.

It is hoped that this might be a fruitful tool for policy development, whereby a path (there may be several) between a nation's current economy and one they wish to emulate identifies a sequence of practicable, achievable structural changes that result in the desired economic improvement. Or it might provide evidence that such improvement is unachievable.

The basic object of study is an Input-Output table. The Input-Output tables that form the bulk of the data for the analysis were from the STAN collection of Input-Output tables maintained by the OECD. Each table presents the flows of money between sectors, the value-added by each sector in the course of its production, and the final consumption of each sector's production; values are in basic-prices, which is to say that consumption taxes and taxes on imports have been removed.

We begin the paper by discussing value-added and how it can be assessed and decomposed into sector-by-sector contributions using Input-Output tables - similar to work done by the OECD in measuring the value-added in international trade (OECD and WTO, 2012). From this we construct a number of distance measures between Input-Output tables and construct the Economy Space. Following that we consider the problem of developing industry-based economic policy to encourage New Zealand to be more similar to Finland, using the similarity measures used in the construction of Economy Space, and apply Economy Space analysis to this.

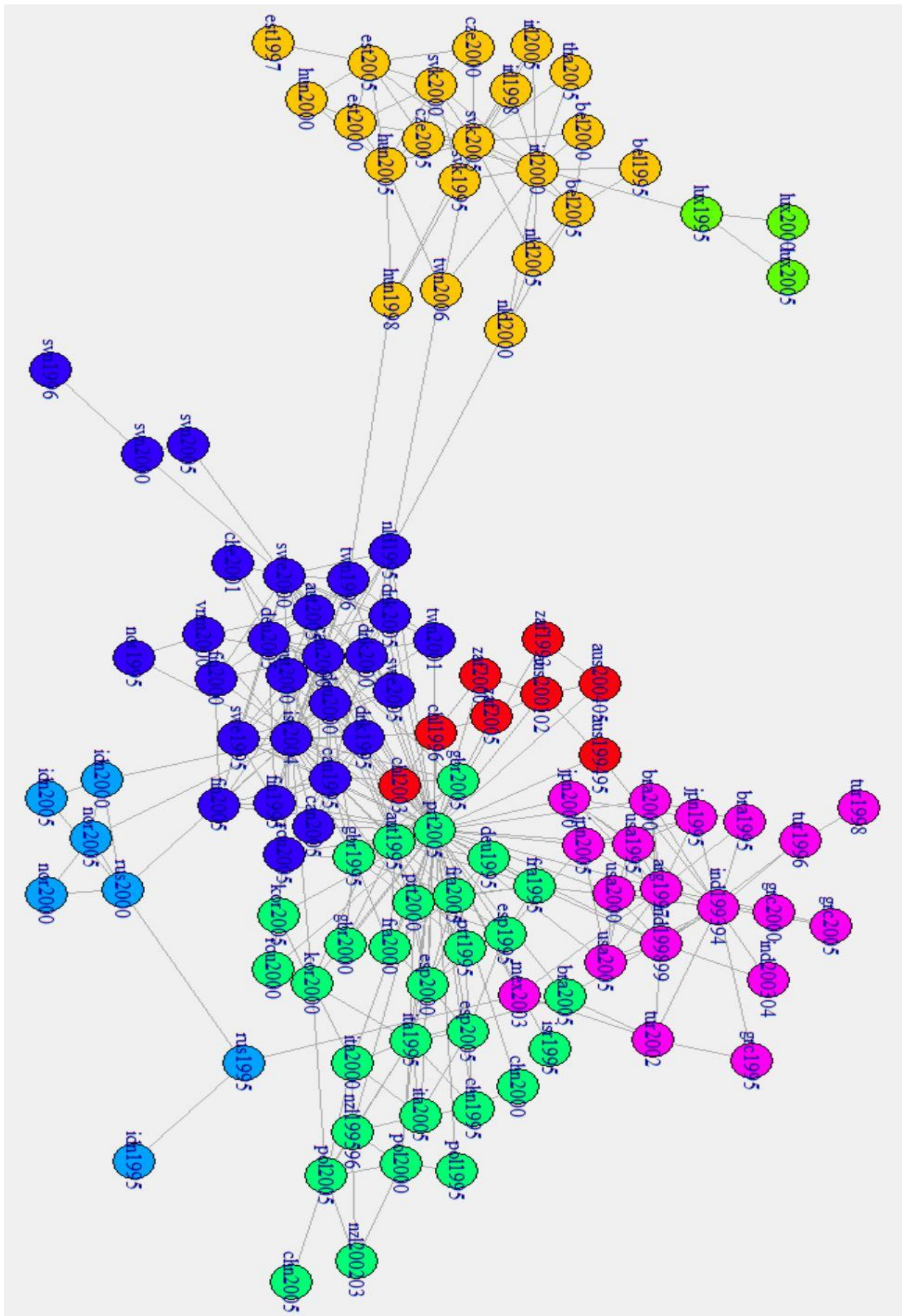


Figure 1: Economy Space: a graph whose vertices are the 115 Input-Output tables in the OECD's STAN repository. The tables are grouped according to similarity and colour coded.

Structural matrices and Gross Value Added

Suppose that an economy consists of n sectors, each providing inputs to other sectors' production as well as providing final production for use or export. Let x be the vector of total output, with x_i being the output by sector i. Note that the inputs to industry i to produce x_i may include imports as well as the input from the n sectors.

Let Z be an $n \times n$ matrix so that the entry Z_{ij} is the amount of input provided by industry i in the production by industry j. From the vector of total outputs and Z we define the technical matrix A as

$$A_{ij} = \frac{Z_{ij}}{x_j}$$

It is clear that A_{ij} is the amount of input required from industry i to produce a unit of output by industry j. Note that unless imports are only non-competitive imports, the columns of A should not be viewed as production functions.

Classically, the technical matrix is related to final use and total outputs as:

$$x = Ax + f$$

sum of use by private house-holds, government, gross fixed capital formation, and exports.

Total output is the sum of value-added and purchases (both imports and domestic). That is,

$$x = v + p + m$$

Where v is the vector of value-added, m is the vector of imports and p the total purchases.

Purchases can be seen to be the column sum

$$p_j = \sum_i Z_{ij} = \sum_i A_{ij} x_j$$

So that (abusing notation by letting the i-th entry of the vector $\frac{a}{b}$ be $\frac{a_i}{b_i}$)

$$\frac{p_j}{x_j} = \sum_i A_{ij} = \sum_i A_{ij} \frac{p_i + m_i + v_i}{x_i} = \left(A^t \frac{p + m + v}{x}\right)_j$$

Or, equivalently,

$$(I - A^t)\frac{p}{x} = A^t \frac{m + v}{x}$$

Hence, when $(I - A^t)^{-1}$ exists, we have that a unit vector of outputs decomposes as:

$$1 = \frac{v + m}{x} + \frac{p}{x} = (I + (I - A^t)^{-1} A^t)\left(\frac{v + m}{x}\right)$$

$$= (I - A^t)^{-1}\left(\frac{v + m}{x}\right)$$

Note that A is a non-negative matrix whose column sums are each less than or equal to 1. So by the Peron-Frobenius theorem, A has a non-negative eigenvector y with eigenvalue $\lambda > 0$ providing an upper bound to the spectral radius of A. Scale y so that $\sum_i y_i = 1$; it follows that $\lambda = \sum_i y_i \lambda = \sum_{ij} A_{ij} y_j \leq \sum_j y_j = 1$. Since A and A^t have the same eigenvalues, it follows that the spectral radius of A^t is strictly bounded by 1 when every industry either has non-zero value added or there are imports in the industry's inputs - this is generically the case; hence $I - A^t$ is invertible. This matrix should be compared with the Leontief inverse, which gives the amount of input required to produce an additional unit of final use. $(I - A^t)^{-1}$ gives the amount of value added and imports needed to produce a unit of output.

The inverse can be seen to be the formal sum

$$(I - A^t)^{-1} = I + A^t + (A^t)^2 + (A^t)^3 + \cdots$$

The ij-th entry of $(A^t)^2$ is $\sum A_{ki}A_{jk}$, which can be seen to be the amount of input by industry j needed to produce a unit of production by industry i, where the input from j goes via production in all the intermediate industries. Similarly, the higher powers of A give the proportion of input from a sector into the output of another sector with more and more sectors acting as intermediaries.

Form the matrix

$$C = (I - A^t)^{-1}\Delta\left(\frac{v}{x}\right)$$

Where $\Delta(y)$ is the diagonal matrix with the ith diagonal entry equal to y_i, for a vector y.

The analysis above implies that the rows sums of C are each less than or equal to 1. The entries of C are non-negative, and hence every entry is less than 1. Moreover, the *i*th row of C gives the breakdown of the value-added in a unit of output from industry *i* into value-added amounts from each of the industries.

The New Zealand economy in 2007

In 2012 NZ published an Input-Output table for the year ending March 31st, 2007 (Statistics New Zealand, 2012). The matrix of contributions to value-added can be visualized as a network graph, where the vertices are sectors and there is a directed edge between sector *a* and sector *b* if a sufficient contribution (as a proportion) by sector *a* is made to the final production from sector *b*. We have chosen to require a contribution of at least 0.01 in order for an edge to exist. The graph is highly connected, in part due to the ubiquity of input from the Banking and Financial Services sector, and so there is little value in displaying the full graph. Looking at a single sector at a time, we can plot the sectors that provide input into that sector's output, and what sectors that sector provides input into: the chain of value-added within the economy that involves that sector

Figures 2 and 3 illustrate for the sector 'Oil and gas extraction'.

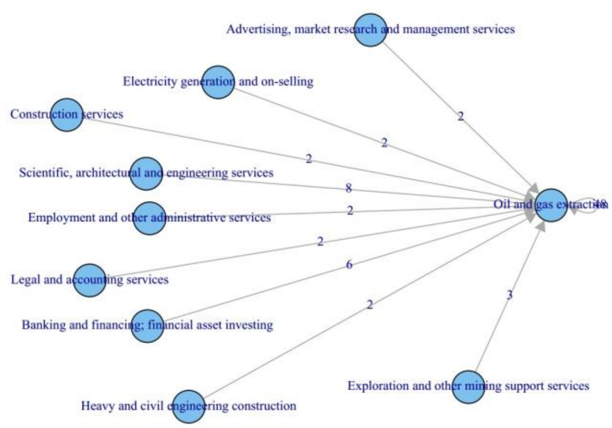


Figure 2: This graph depicts the industries that provide significant proportions of value-added to the total value-added by the Oil and Gas Extraction sector. The number on an edge pointing away from a sector is the percentage of total value-added that is provided by that sector.

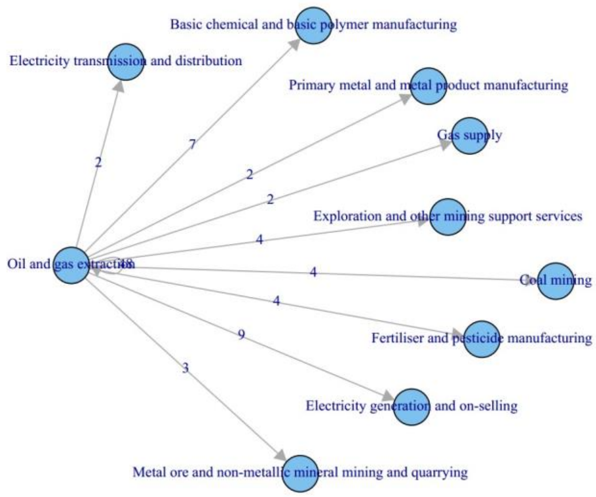


Figure 3: As in figure 2 this graph depicts what industries the Oil and Gas Extraction sector provides significant value-added to. The number on an edge pointing to a sector is the proportion of that sector's value-added that is provided by the Oil and Gas Extraction sector.

Decomposing economic indicators

Continuing with our construction, let f be the vector of final demand (exports, domestic consumption, capital formation and changes in stock provided by each industry). The Gross Value Added (as a vector with an entry for each industry) is

$$GVA = C^t f$$

Gross Domestic Product is obtained from Gross Value Added by the addition of taxes on imports, value-added tax, and taxes on production.

An economic indicator is a number associated to an economy at a given time that is meant to provide information about certain aspects of the economy. For better or worse, these economic indicators are used to compare one economy to another, or to formulate goals for economic development, say in terms of growth in real GDP or the Gini coefficient. Yet two economies can have similar economic indicators yet be very different structurally. In fact, for any set of economic indicators, it is mathematically possible for two economies to have exactly the same values for these indicators, yet not be structurally equivalent, for whatever sensible definition of structural equivalence you might consider. Economic indicators are often statistics of a random variable. Though the value of the statistic can be useful, much more information is encoded in the distribution of the random variable. Unfortunately, these distributions may not be known or only infrequently known (with the latter case making it of little value in making comparisons between economies). For a number of economic indicators, Input-Output tables provide a view on the distribution of the data that underpins the indicator – generally, economic indicators that relate to production and consumption.

Example: exports as a percentage of GDP

New Zealand economic development policy has since 2012 been driven by the government's Business Growth Agenda (Ministry of Business, Innovation, and Employment, 2017). The Business Growth Agenda addresses a number of different areas, such as Infrastructure, Investment, and Export Markets (to name three of the six areas). The Export Markets area sets aspirational targets for exports expressed as a percentage of GDP: 40%. To give some context, New Zealand has had exports approximately 30% of GDP for the last 40 years – it is likely that the issue is a structural one and not just a matter of increasing export volumes or prices.

On the face of it, increasing exports also increases GDP, so exports as a percentage of GDP should be asymptotic to the reciprocal of the proportion of domestic inputs used in producing exports – this is a restatement of the expression of GDP as a sum of exports, investment, and domestic use less imports - but there could be a multiplier effect that increases GDP beyond the increase due to net exports, reducing the ratio.

Thus, the role of other industries in exports production and the non-tradable sector is important for understanding how exports can be increased as a proportion of GDP.

For a given Input-Output table in basic prices, we can consider for each industry the ratio between the exports done by that industry and the total domestic value-added in the production of that industry. That is, we can measure the 'exports as a percentage of GVA' at an industry level. The national exports as a percentage of GVA indicator is obtained from the industry-level figures as a weighted average.

Order the industries in some manner – later we will stipulate a particular ordering. We perform a thought experiment where we suppose that an economy consists of the first k industries in the ordering plus only the inputs required from the excluded industries to produce what these k industries produced for final use in the year that the Input-Output table was reporting on.

The remainder of the production by the $n - k$ industries is then assumed to be substituted by imports - this is clearly an unrealistic setup as production by non-exporting industries will be non-tradable, but we are simply presenting a heuristic. Suppose that all the industries in the economy have positive value-added.

With this set-up, at the k-th step the economy has a final use vector f^k where $f_i^k = f_i$ when $i \leq k$ and zero otherwise. The output required to produce f^k is given by $x^k = (I - A)^{-1} f^k$, and the value-added is

$$v^k = \Delta\left(\frac{v}{x}\right)(I - A)^{-1} f^k = C^t f^k$$

Let e_i be the amount of exports done by industry i, and let g_i be the total value-added in the production of x_i, so that $g_i = x_i \sum_j C_{ij}$. (Note this value-added measure includes value-added by industries other than i, different from $C^t f$; in particular the sum $\sum g_i$ will be much larger than GVA due to significant double counting).

For industry *i* we can consider the ratio $\frac{e_i}{g_i}$ which compares the amount of exports to the total value added to the economy by the industry's production. If this is greater than 1 then imports play a significant role in production; if it is much smaller than 1 then exports aren't a significant part of final consumption or imports play a small part in production. Renumber the industries so that industry 1 has the largest value of $\frac{e_i}{g_i}$ and industry *n* the least; in the case of ties, subsequently order on decreasing values of industry value added, g_i.

We compute a value

$$r_k = \frac{\sum_{i=1}^{k} e_i}{\sum_{i=1}^{n} v_i^k}$$

$k = 1,...,n$, which is a non-increasing sequence (because of the order we imposed on the industries - with a different ordering it wouldn't necessarily be non-increasing) with r_n equal to the ratio of exports to GVA. The denominator of r_k is the value-added in the production of the first k industries that is not contributing to production by the remaining industries. These k sectors correspond to a proportion p_k of GVA, so we have constructed n pairs of the form (p_k, r_k); because p_k is an increasing sequence we can suppose that these points are samples from a continuous function on $(0,1]$. Choosing a suitable method of interpolation provides an estimate of this function, providing a mapping that takes an Input-Output table and outputs a continuous function on $(0,1]$. The figure below illustrates for the New Zealand Input-Output table for 2007.

Structural Curves

We can repeat that construction for a number of other economic indicators, all that is needed is the ability to compute these indicators on subsets of the set of sectors. In particular, we can repeat the construction four times as described now:

1. *'The exports profile curve'* This, described in detail above, is a smooth interpolation of the points of the form $\{(p_i, r_i)\}$, where p_i is the proportion of Gross Value Added in the first i industries and r_i is the sum of the exports done by the first i industries divided by the Gross Value Added by those industries' production (including the necessary value-added inputs from the sectors not in those i).

2. *'The domestic involvement curve'* With this curve, the intention is to measure how much of non-imported value-added in production is sourced from the producing industry. The notion is that for industries where this number is small, the impact on changes to production has wider impact on other industries than for industries with a larger number. Industries with high values are 'decoupled' from the rest of the economy.

   The curve is a smooth interpolation of the points of the form $\{(p_i, t_i)\}$, where p_i is the proportion of Gross Value Added in the first i industries, as above; and t_i is the weighted mean

$$t_i = \frac{\sum_{j=1}^{i} v_j^i \frac{C_{jj} x_j}{g_j}}{\sum_{j=1}^{i} v_j^i}$$

   The weighting is done to give more emphasis to industries whose production provides a greater contribution to GVA.

3. *The Exports participation curve* This curve is obtained as a smooth interpolation of the points $\{(p_i, s_i)\}$ where p_i is the proportion of Gross Value Added present in the first i industries and s_i is the proportion of exports present in the first i industries. Analogous to the Gini curve, how far this curve is from the line $s = p$ indicates how concentrated exports are amongst all the sectors.

4. *Labour's share of value-added curve* A significant component of value-added is the contributions of labour, and this curve attempts to measure the importance of labour to production. Let u_i denote the

proportion of value-added in production of salary and wages paid by sector i in its annual production. A function on (0,1] is obtained as an interpolation of points $\{(p_i, z_i)\}$ where p_i is the proportion of GVA present in the first i sectors and z_i is

$$z_i = \frac{\sum_{k=1}^{i} v_k^i u_k}{\sum_{k=1}^{i} v_k^i}$$

The graph of this function is the *Labour's share of value-added* curve.

Figure 4 plots unsmoothed versions of these curves for the Input-Output table for New Zealand 2007.

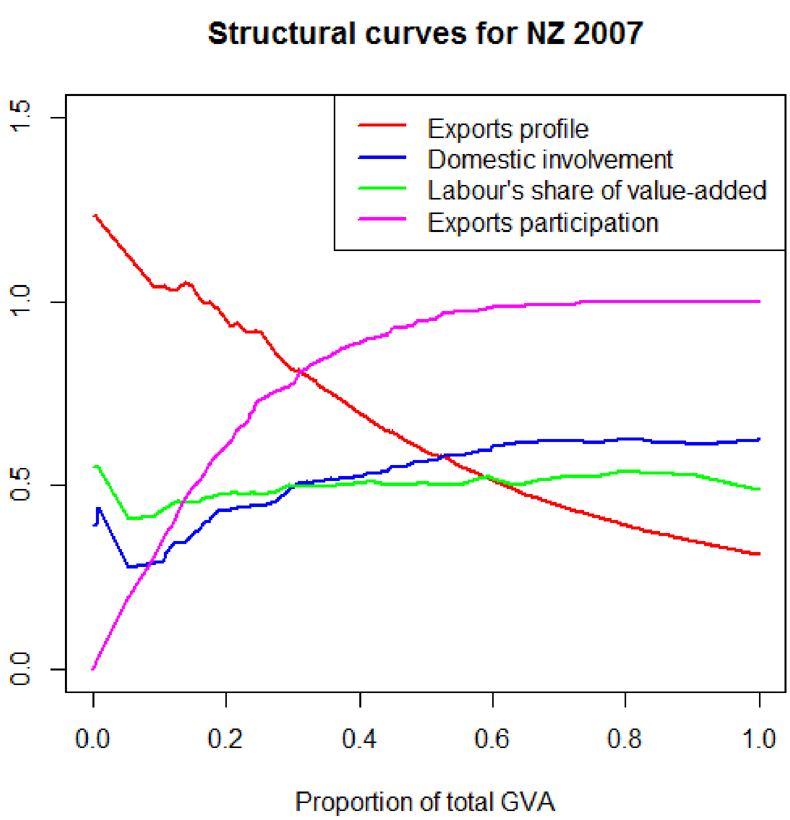


Figure 4: The four curves here are unsmoothed versions of the four curves described above. The *Exports participation curve* corresponds to *Proportion of*

total exports; the *Exports profile curve* corresponds to *Exports to GVA*; the *Labour's share of value-added curve* corresponds to *Employee compensation to GVA*; and the *Domestic involvement curve* corresponds to *Proportion of direct value-added to GVA*.

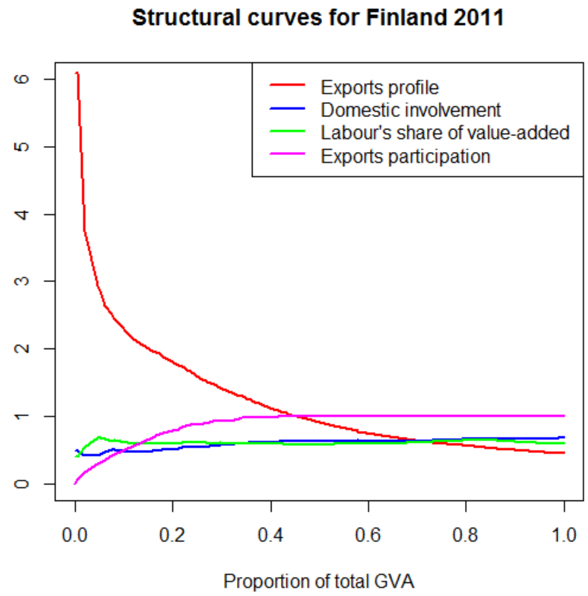


Figure 5: Structural curves for Finland 2011, using the same axes as those used for the plot of New Zealand 2007.

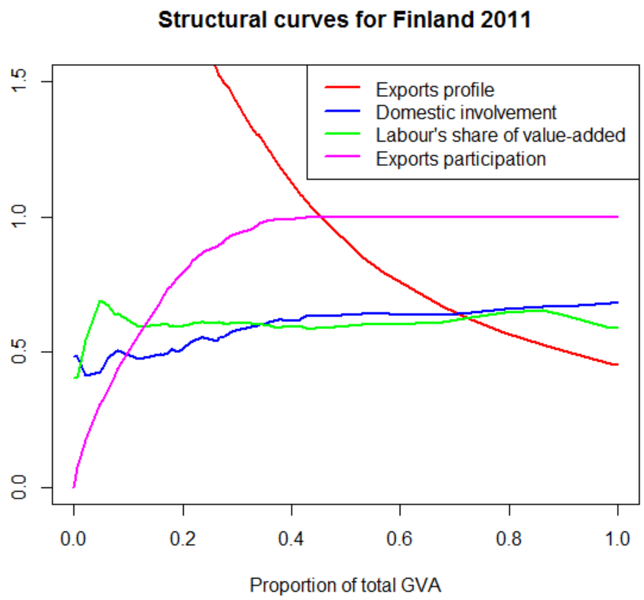


Figure 6: A replot of Finland 2007 using axes that allow the shape of all the curves to be seen.

As another example, figures 5 and 6 present the same curves for Finland 2011 (Statistics Finland, 2017). The *Exports to GVA* curve for Finland 2011 has a larger maximum than the same curve for New Zealand 2007 so we plotted the curves twice, one with the same axis scales as in the plot of New Zealand 2007 and another showing all of the *Exports to GVA* curve.

The Input-Output tables in the STAN collection have been harmonized to consist of at most 48 sectors, and each sector definition is the same throughout the collection. This has been done by taking more granular Input-Output tables and aggregating appropriate sectors. Thus, any curve we construct on the STAN Input-Output table would be seen as a smoothing of the curve based on the more granular Input-Output table. In practice, the values near 0 might be very different on the two functions, but they will be more and more similar as we approach 1. But there is a concern in comparing two graphs that are based on

Input-Output tables that have a different set of sector definitions. Nevertheless, when the sector definitions are the same the mapping of Input-Output tables to continuous functions on (0,1] provides a meaningful way to measure the distance between two Input-Output tables.

Distance measures on Input-Output tables

There are 115 Input-Output tables in the STAN database release for 2006 (later releases contain more recent tables but are for a restricted set of countries), representing 44 economies in the OECD. There were four Input-Output tables that omitted information about salary and wages, making it impossible to compute Labour's share of value-added. For those cases (Argentina 1997, India (1993-94,1998-99) and Israel 2004), we set all distances arising from Labour's share of value-added equal to zero. Let T denote the set of 115 Input-Output tables in the STAN collection.

For each Input-Output table *x* we construct four functions

$$f_1^x, f_2^x, f_3^x, f_4^x: (0,1] \longrightarrow [0,\infty)$$

where these functions are the four functions described in the previous section constructed using the table *x*.

These functions provide distance functions

$$d_i: T \times T \longrightarrow [0,\infty)$$

Where

$$d_i(x, y) = \int_0^1 \left|f_i^x(t) - f_i^y(t)\right| dt$$

Let N_i be the standard-deviation of the set of values obtained by applying d_i to the set of distinct unordered pairs from T ×T. We then obtain a distance function $\widetilde{D}: T \times T \longrightarrow [0, \infty)$ as a weighted sum

$$\widetilde{D} = \frac{d_1}{N_1} + \frac{d_2}{N_2} + \frac{d_3}{N_3} + \frac{d_4}{N_4}$$

Finally, to deal with dependencies among the $d_1, \ldots, d_4$, let N be the standard-deviation of the set of values obtained by applying $\widetilde{D}$ to the set of distinct unordered pairs from T × T and define

$$D = \frac{1}{N} \widetilde{D}$$

D is a distance function on the set of Input-Output tables for which all the four economic indicator curves can be calculated, not just those in T.

'Economy space'

We can form a weighted undirected complete graph G whose vertices are T and where the weight on an edge is given by the distance function D. Some distances are too large to justify the existence of an edge, so we develop a criterion for removing edges based on the edge weight relative to the weights on the co-terminal edges, such that the resulting graph is connected.

For any graph Γ we denote by $V(\Gamma)$ the vertices of Γ, and $E(\Gamma)$ the edges of Γ. For an edge e, the weight attached to e is $w(e)$.

Let S be a minimal spanning tree for G. For each vertex v in S let w_v be the minimum of the weights of the edges in S containing v. For a small number $\epsilon > 0$, we define a set of edges $\mathcal{E}_\epsilon$ in G as

$$\mathcal{E}_\epsilon = \bigcup_{v \in V(S)} \{e \in E(G) |\ v \in V(e)\ \wedge |w(e) - w_v| \leq \epsilon\}$$

and then define a graph G to be the result of adding $\mathcal{E}_\epsilon$ to S.

We want to choose ϵ so that there is strong evidence that the vertices in G are strongly clustered. A vertex clustering is simply an assignment of labels to vertices; vertices with the same label are in the same cluster. Another common terminology is to call a cluster a *community* and techniques for assigning labels to vertices in graphs are *community detection* methods. For our purpose a useful clustering is one where intra-cluster links are more common than inter-cluster links, and we use *modularity,* due to Newman (Newman, 2006), to provide a measure of this.

Consider a clustering C where a vertex v is labelled n_v with n_v taking values in $1, \dots, L$, then the modularity of the clustering is defined as

$$mod(C) = \sum_{i=1}^{L} (e_i - a_i^2)$$

Where e_i is the proportion of the total edge weights (counting undirected edges twice) contributed by those edges joining vertices in cluster i; if the graph were unweighted, it would be the observed probability that an edge joins vertices in cluster i. The value a_i is the proportion of the total edge weights (counting undirected edges twice) of the edges with a vertex in cluster i; for unweighted graphs it is the probability that an edge is incident to the vertices in cluster i. If the clustering were entirely random, and edge formation were independent of the clustering, then the value e_i should equal a_i^2. The difference between e_i an a_i^2 provides evidence that the connectivity within cluster i is more than expected - and hence the connectivity between cluster i and any other cluster is less than expected. Hence $mod(\mathrm{C})$ is large and positive when strongly weighted edges connect vertices in the same cluster more than the average, and large and negative when strongly weighted edges connect edges in different clusters more than the average.

When modifying an existing clustering to increase the number of clusters, we see that generally both e_i and a_i would be expected to decrease, so whether the modularity improves depends on whether a_i^2 decreases more than e_i does.

In our situation, we are not looking only at different communities on a single graph but rather community detection algorithms applied to a nested sequence of graphs. Our challenge is to choose a graph where there is strong evidence of clustering; however, modularity is affected by the number of edges, so we need to account for that.

We applied four different community detection algorithms to the sequence of graphs obtaining by varying ϵ between 0 and 0.5, namely *the Multi-level method* (Blondel, Guillaume, & Lambiotte, 2008), the *Walk-Trap method* (Latapy., 2006), the *Spinglass* method (Bornholdt, 2006), and the *Infomap* method (Bergstrom, 2008). These algorithms were all implemented in the computing environment R within the *igraph* package (Gabor Csardi and Tamas Nepusz. InterJournal, 2006) (R Core Team, 2015). In each case, we applied the algorithms with the weight for an edge with distance w_0 being

$$prob(w_e \leq w_0 | e \in E(G))$$

- this is so that low-distance is translated into high weight, and vice versa.

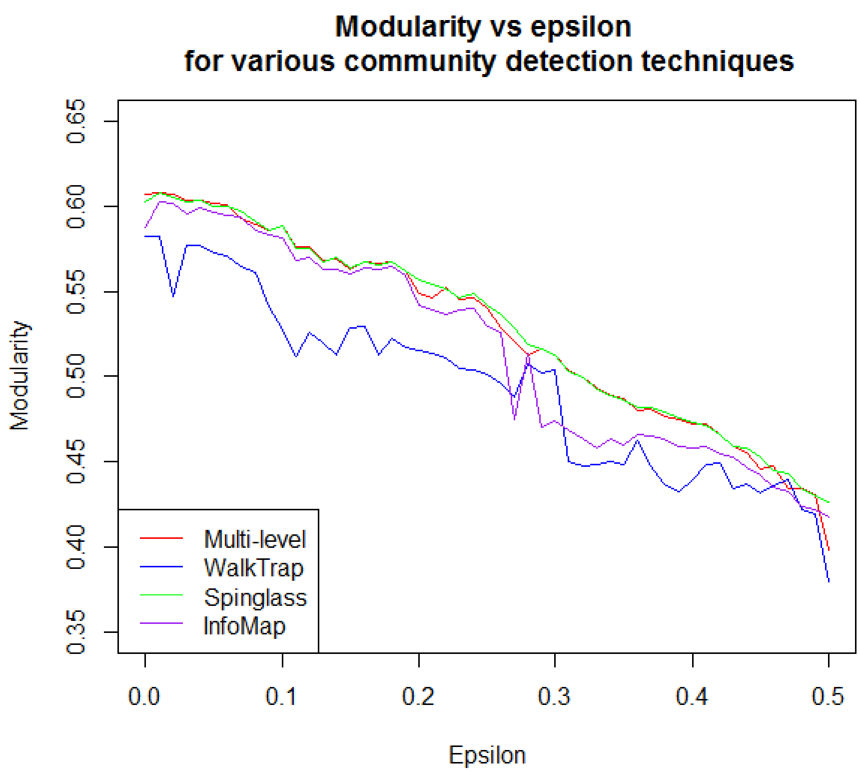


Figure 7: How modularity varies with ϵ, for the four community detection methods.

The graph 7 illustrates how the modularity varies with ϵ with each of these approaches. Note that the number of communities varies in each method as ϵ varies. Both the Multi-level and Spinglass approaches are variations on maximum modularity approaches, and so as you might expect they perform quite similarly. The WalkTrap method provides clustering with consistently lower modularity than the other three. The Infomap method lies somewhere in between, though fairly similar to the modularity-based approaches for moderate ϵ. All the approaches give their highest modularity at $\epsilon = 0.01$, on between 8 and 14 communities, though for the Multi-Level and Spinglass methods the modularity varied slowly for small values of ϵ, so there is some argument that small decreases in modularity for commensurately larger increases in graph complexity justify selecting a larger value of ϵ.

Modularity on its own is not sufficient to select between the Multi-level and Spinglass methods, nor to identify what value of ϵ to select. We further compared the two methods by considering how well the clustering provided a generative model of the graphs.

A clustering provides a set of probabilities for edge existence, namely

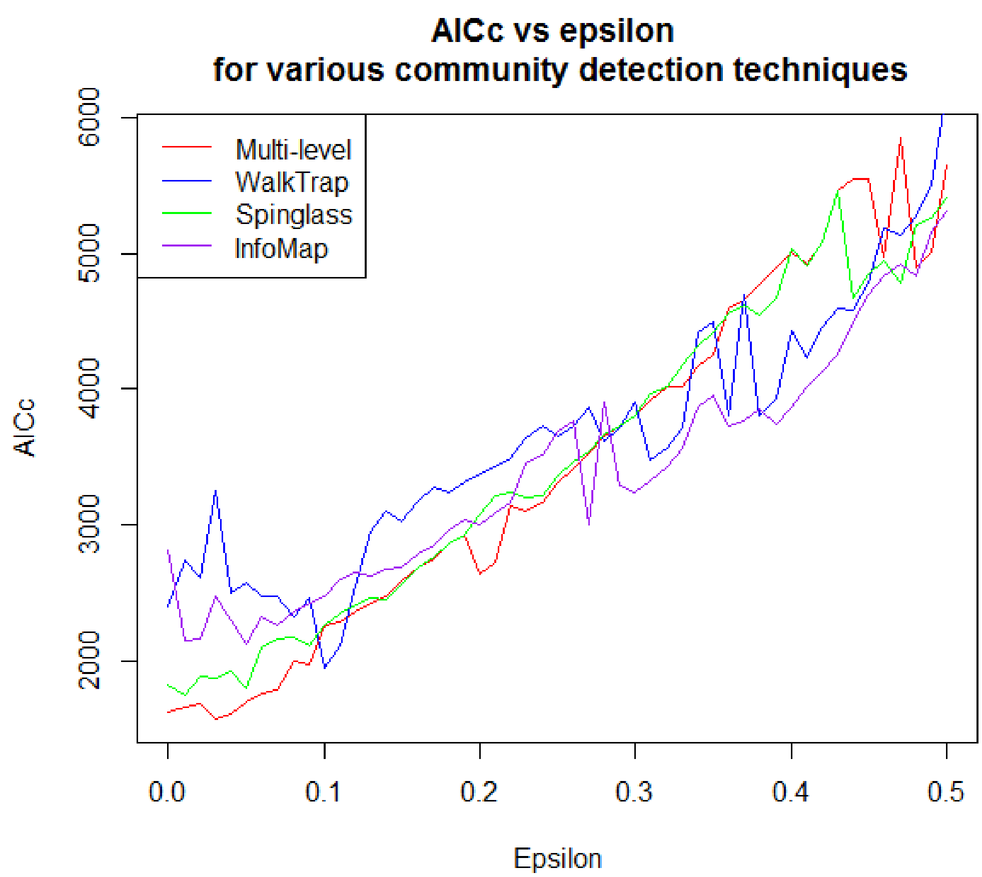
$$prob(e|C) = \frac{|\{(v,w) \in E(G) |\ n_v = n_{e(0)} \wedge n_w = n_{e(1)}\}|}{|E(G)|}$$

That is, the probability of an edge is approximated by the probability that the clusters containing the endpoints of the edge are connected. The log-likelihood of observing the edges can then be computed. As the number of clusters increases, this estimate becomes the estimate that each pair of vertices is connected, which would be the most accurate model of edge prediction obtained in this manner – but at the penalty of a large number of parameters. We penalize the log-likelihood by applying the Akaike Information Criterion, adjusted for small amounts of data. More precisely, we computed:

$$AICc(G,C) = -2loglike(G;C) + 2k(C) + \frac{2k(C)(k(C)+1)}{|E(G)| - k(C) - 1}$$

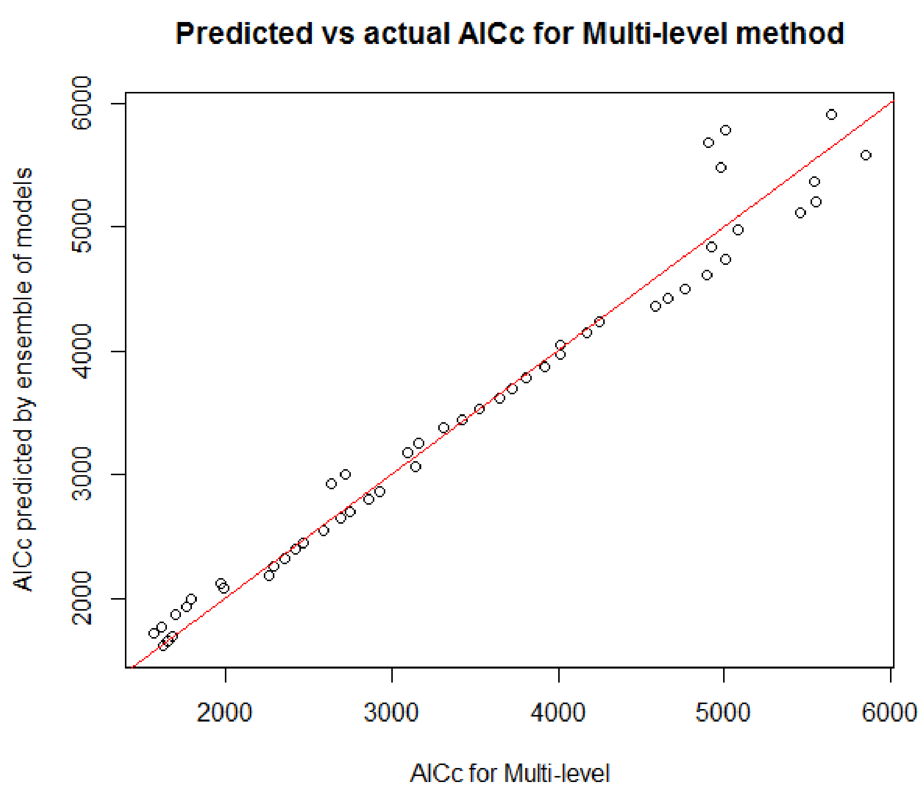
Where $k(C) = L^2 - 1$ when the clustering C has L clusters. Smaller values of $AICc$ are indicators of better fit than larger values, for a fixed G.

The figure below shows how the four clustering methods compared.

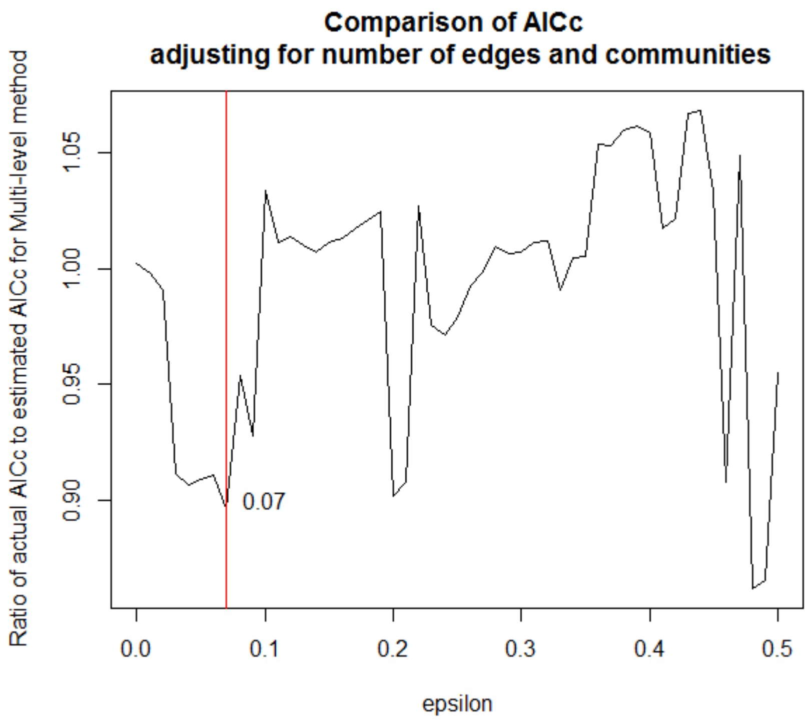


For a fixed vertex set and clustering, $AICc$ is increasing with the addition of edges. Heuristically, we expect $AICc$ to increase with the number of clusters, everything else being equal. Hence in order to use $AICc$ to measure how well the communities model the graphs, we want to account for the increases due to increased numbers of edges or changes due to changes in the number of clusters. We randomly generated 10000 distinct samples of size 40 from $\{0,0.01,\ldots,0.5\}$ and for each sample built a linear regression of $AICc$ for the Multi-Level approach against the number of edges and the number of clusters. In general, these models explained $AICc$ extremely well, with large adjusted R-squared values exceeding 0.99.

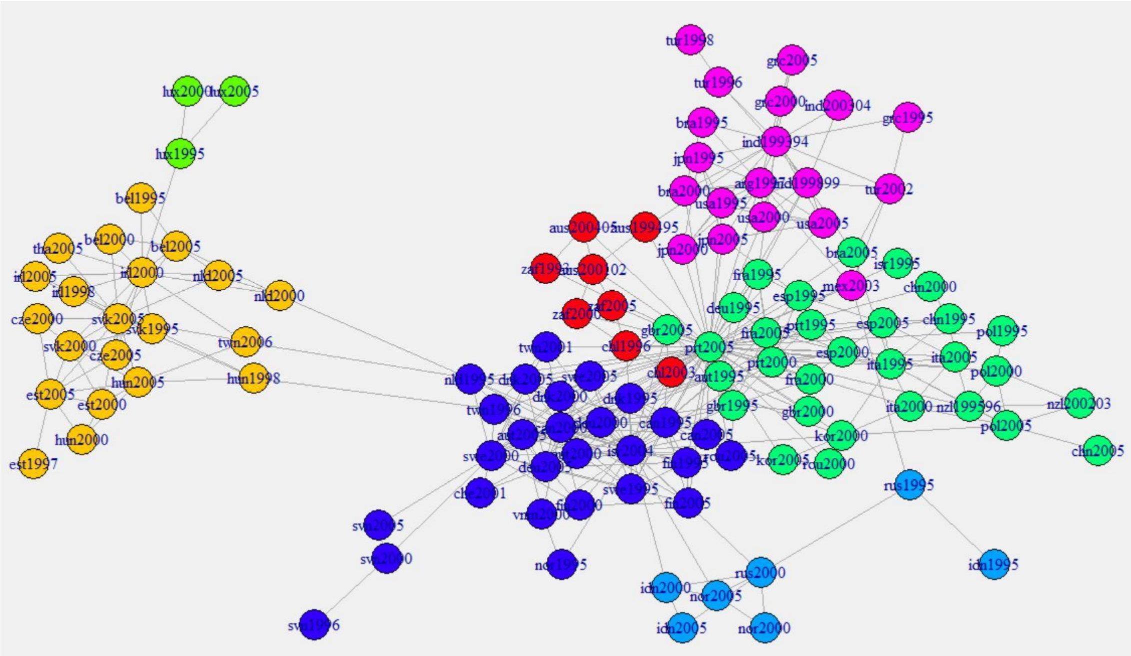
We created an ensemble model by weighting each linear regression by the adjusted R-squared less 0.99, and used this model to predict values of $AICc$ as a function of ϵ. The graph below illustrates how the prediction compared to the actual values.



Plotting the quotient of actual $AICc$ and predicted $AICc$ against ϵ shows the instances when actual $AICc$ is smaller than predicted, which we take as evidence that the graph structure is better modelled.



We find that for $\epsilon = 0.07$ the Multi-level method of community detection provides its optimal generative model of the graph. There are 7 communities on a graph with 115 vertices and 388 edges.



An application to economic policy development

In this section, we consider the New Zealand economy as represented by its 2013 national input-output table (the most recent at the time of writing) (Statistics New Zealand, 2016), and use the Economy Space graph to identify a sequence of changes that would make the New Zealand economy more similar to that of Finland in 2011 (using the similarity measures developed previously). The choice of Finland is arbitrary, though it is not uncommon to compare New Zealand to Finland, as Finland is also a small, developed nation.

It is worth remarking again that the choice of measures has a strong effect on the topology of Economy Space, and hence on the nature of any economic policy created to enable a decrease (or indeed, an increase) in distance between one economy and another. Here we chose exports-related measures, notably exports in proportion to GVA, so any suggested strategy will be one that is intended to affect those measures, regardless of whether they are either feasible or goals that would result in socio-economy benefit. Setting targets for key statistics (such as exports in proportion to

GDP) can have unintended consequences, so Economy Space analysis might prove useful in understanding what such consequences might be, before these targets are set. The approach is to convert the two input-output tables to the industry classification used in the STAN tables, update the Economy Space graph to include these tables, and then analyse the shortest paths between them. The analysis of these paths will be rudimentary, and far short of the nuanced analysis needed for evidence-based policy; however, it will illustrate the usefulness of the approach, and raise a number of questions and hypotheses that would naturally be addressed in a more in-depth treatment.

A key challenge is translating both the Finnish industry classification and the New Zealand industry classification to the STAN industry classification in such a manner that the input-output tables can be converted with minimal error. This involves establishing a concordance between industry classifications, where a partial match is in accord with how much production is allocated.

The New Zealand 2013 input-output table is in terms of 106 industry categories and there is a concordance with the Australia-New Zealand Standard Industry Classification of 2006 (ANZSIC06), which has 506 categories at level 4. The STAN industry classification has 48 industries, and there is a concordance to ISIC Rev 3.1; the concordance between ISIC Rev 3.1 and ANZSIC06 provides a common expression for both the STAN classification and the NZ IO classification.

Concretely, this expression for the New Zealand IO industry classification is a matrix, which we denote T_1, with a column for each of the 106 industries, and where the entries in each column are the proportions of the 506 ANZSIC06 categories that make up the respective industry. The rows of T_1 each sum to one. Note that the concordances do not (and cannot) specify the proportions when they are not zero or one, merely that the contribution is ‘partial’. Part of our task is to estimate these proportions when they are not zero or one, so as to minimize the error in our conversions.

Similarly, we have a matrix T_2 which provides the encoding of the STAN IOs in terms of the ANZSIC06 classification; T_2 is a matrix with 506 rows and 48 columns, whose rows each sum to one. Again, we will need to find estimates for the entries of T_2 that are not zero or one.

Suppose that for an NZ IO industry category i and STAN IO industry category a, we let B_{ai} be the proportion of category i that is category a. If Z is a transactions matrix in terms of the NZ IO categories, then we obtain the input by STAN category a into production by STAN category b as $\sum_{ij} B_{ai} Z_{ij} B_{bj}$ – the first factor in each summand represents the combinations of rows that make up a and the second factor the combination of columns that make up b. Thus, as a matrix, the translation of Z to a transactions matrix in terms of STAN IO is BZB^t. Thus we need to find the matrix B, which is a akin to a change of basis matrix; the columns of B are the ‘coordinates’ of the NZ IO categories in terms of the STAN IO categories. Note that each of the columns of B sum to one.

Assuming the specification of T_1 and T_2 for the moment, the natural linear map between quantities expressed in terms of NZ IO categories and quantities expressed in STAN IO categories is a projection. If v is a column of T_1 and the columns of T_2 are $w_1, \ldots, w_{48}$, then the projection of v into the span of $\{w_1, \ldots, w_{48}\}$ is (in coordinates)

$$v \mapsto (v \cdot \frac{w_1}{||w_1||^2}, \ldots, v \cdot \frac{w_{48}}{||w_{48}||^2})$$

Here, the norm of a vector, $||x||$, is the square root of the sum of the squared entries of x. This projection is unsuitable as it is not the identity on the span of $\{w_1, \ldots, w_{48}\}$ as the columns of T_2 are not orthogonal. Suppose instead that there are constants $a_1, \ldots, a_{48}$ so that the map is of the form:

$$v \mapsto \left(v \cdot \frac{w_1}{a_1}, \ldots, v \cdot \frac{w_{48}}{a_{48}}\right)$$

We require that passing between the two classification should preserve the total apportionment of ANZSIC06 classes; this is to say that

$$\sum_{i,j,k} \frac{(T_1)_{ki}(T_2)_{kj}(T_2)_{pj}}{a_j} = 1 \ \forall p$$

We find that the equation holds with $a_j = \sum_s (T_2)_{sj}$, which is to say that the a_j are the column sums of T_2.

With this, define the following matrices:

$$L_1 = T_2^t T_1 \Delta\left(\frac{1}{C(T_1)}\right)$$

$$L_2 = T_1^t T_2 \Delta\left(\frac{1}{C(T_2)}\right)$$

Where $C(X)$ is the vector of column sums of X.

Note that the column sums are one for both L_1 and L_2, and play the role of the matrix B discussed above.

If Z is the transactions table for NZ 2013 in terms of the NZ IO classification, the transaction table in terms of the STAN IO classification is

$$Z_0 = L_1 Z L_1^t$$

We observe that the sum of the entries in Z_0 is the same as the sum of the entries in Z because the column sums of L_1 are all one.

The conversion back to a classification in terms of NZ IO categories is

$$Z_1 = L_2 Z_0 L_2^t$$

We estimate the entries of T_1 and T_2 that are not zero or one by minimizing the difference between the entries in Z_1 and the entries in Z. L_1 is a 48-by-106 matrix, with 103 entries that are uncertain, distributed over 36 columns; L_2 is a 106-by-48 matrix, with 256 uncertain entries distributed over 30 columns. Thus, in total there are 293 unknowns on $106^2 = 11236$ equations, which is an extremely over-determined problem. In practice, we were not able to solve the problem exactly, but used the following approach:

1. Minimize the sum of the squares $\sum_{ij}\left(Z_{ij} - (L_2 L_1 Z L_1^t L_2^t)\right)^2$ to obtain solution θ_0. Let the corresponding Z_1 be denoted $Z_1(\theta_0)$.
2. Order the columns of $Z_1(\theta_0)$ in decreasing order of absolute error in the column sums, as $n_1, \ldots, n_{106}$. Now iterate as follows:
    a. Minimize the absolute error in the sum of column n_1, using numeric methods and starting at θ_0, to obtain solution θ_1.
    b. Having minimized the sum of the absolute errors in the sum of the columns $n_1, \ldots, n_k$ with parameters θ_k, now minimize the sum of the absolute errors in the sums of the columns $n_1, \ldots, n_{k+1}$ using numerical methods starting at θ_k to obtain solution θ_{k+1}.
3. Now order the rows of $Z_1(\theta_{106})$ in decreasing order of the absolute error in the row sums, and repeat Step 2, but minimizing over the absolute error in the column sums plus the sequence of absolute errors in the row sums.

This produces solution θ_{212}, which provides our estimate of $L_1$and L_2.

A similar process is used for converting the Finish input-output table for 2011 to STAN IO categories, only the conversion is via ISIC Rev 4 rather than ANZSIC06.

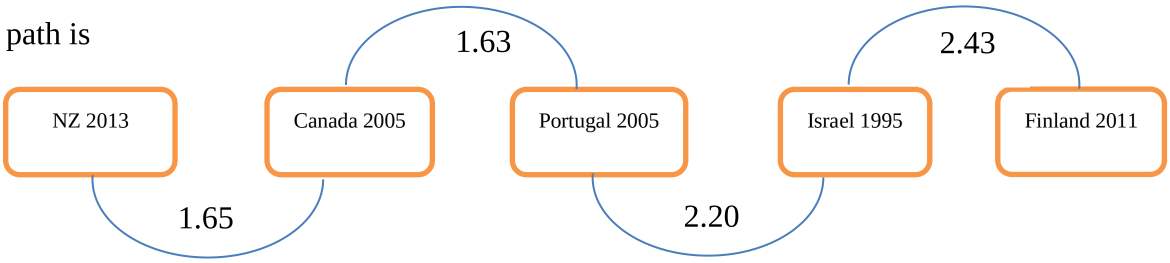
Paths in Economy Space between New Zealand 2013 and Finland 2011

Now that we have represented the New Zealand and Finnish input-output tables in terms of the STAN input-output industry classification, we adjust the Economy Space graph as follows:

1. Add two new vertices, one for New Zealand 2013 and the other for Finland 2011.
2. Compute the distances between each of these vertices and the 115 STAN input-output tables.
3. Compute the distance between New Zealand 2013 and Finland 2011.
4. For each of the two new vertices, create an edge between that vertex and the vertex that is nearest (which might be the other new vertex). Also create an edge between any vertex whose distance is within $\epsilon = 0.07$ of that distance.

We found that this only added 3 edges and 2 vertices to Economy Space. Note that the distance between New Zealand 2013 and Finland 2011 was 3.72, whereas the nearest neighbour to New Zealand 2013 was Canada 2005 (distance 1.65) and the nearest neighbour to Finland 2011 was Israel 1995.

There is a unique shortest path, even when allowing ϵ extra distance per edge. This path is



The shortest path has length 7.91, in contrast to the measured distance of 3.72. However, the points on this path are economies that we know can exist, whereas we do not know that there is a path directly between New Zealand 2013 and Finland 2011 of length 3.72 that consists of realisable economies, or that it is possible to move directly between New Zealand 2013 and Finland 2011 (in, say, a year).

Translating paths in Economy Space into plans for economic development

The distances between input-output tables are computed as the weighted sum of differences between four structural curves. For each input-output table the curves constructed are dependent on an ordering of the industry categories, a different ordering possibly for each input-output table. With the ordering fixed, the curves are obtained as continuous interpolations on a sequence of weighted averages of values for each industry. Thus, the curves are determined by the values of certain measures for each industry, the weights on each industry (generally the industry's value-added), and the ordering on the industries (which is in terms of the ratio of an industry's exports to the domestic input required to produce that industry's output).

To change the curves, we need to know whether the industries need to change their order or need to change their relative contribution to the economy or need to change the value of the measure.

Three sets of industry benchmarks

Recall for an industry i the following quantities derivable from an input-output table: the proportion of value-added that is salary and wages, u_i; exports as a proportion of domestic value-added, $\frac{e_i}{g_i}$; and the proportion of domestic value-added that is contributed by industry i, $\frac{c_{ii}x_i}{g_i}$.

Let F_1 be the cumulative distribution function for $\{\frac{C_{ii}x_i}{g_i}\}$; F_2 the cumulative distribution function for $\{u_i\}$; and F_3 the cumulative distribution function for $\{\frac{e_i}{g_i}\}$. Then we define three benchmarks on pairs of industries i and countries n as

$$c_1^{i,n} = F_1(\frac{C_{ii}x_i}{g_i})$$

$$c_2^{i,n} = F_2(u_i)$$

$$c_3^{i,n} = F_3(\frac{e_i}{g_i})$$

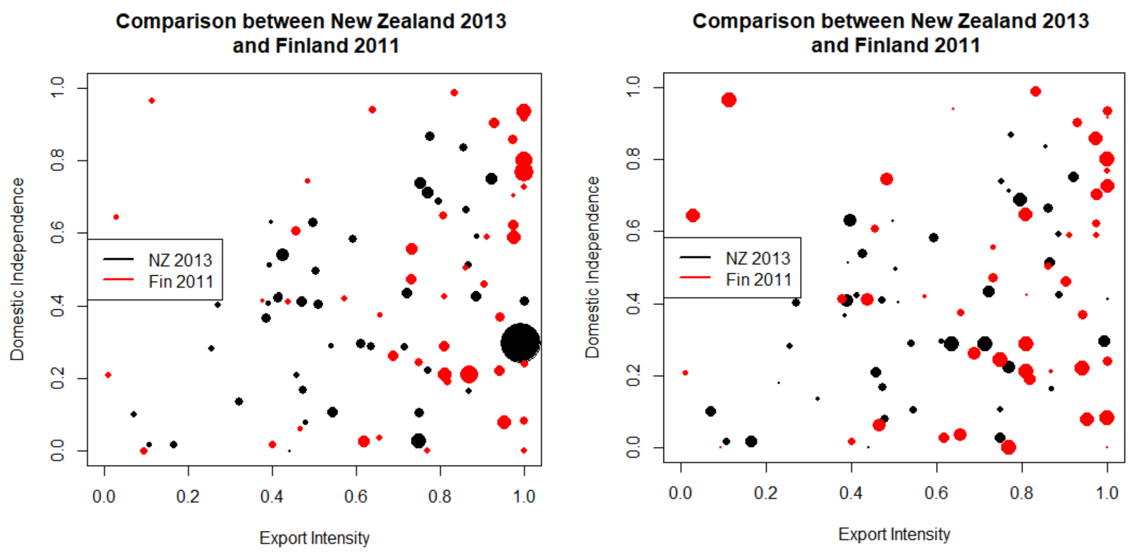
Where it should be understood that the cdf's are being applied to the values for industry i in country n. Each of c_1, c_2, c_3 take values in the unit interval.

We call c_1 the measure of Domestic Independence, c_2 the measure of Labour's Share of Value-Added, and c_3 the measure of Export Intensity.

Comparing economies in the path to Finland

In this section, we use the benchmarks introduced just now to investigate the differences between economies as we move along the path. We develop two graphs for each pair of economies being compared, both plotting the same points, but scaling the size of the plotted point differently. In either case, for each industry in an economy we plot a circle; the radius of the circle is an increasing function of the proportion of the economy's exports done by that industry, in the case on the left-hand side of the page, and proportional to the benchmark value for salary and wages as a proportion of value-added in the graph on the right-hand side; and the coordinates are given by the benchmark values of the industry for two measures: exports as a proportion of domestic value-added, and direct value-added as a proportion of domestic value-added.

The first set of figures show the difference between New Zealand and Finland. In the left-hand graph, we see that New Zealand is less diversified than Finland, with exports dominated by one industry (Food, Beverages, and Tobacco, as it turns out). Moreover, Finnish exporting industries tend to have a higher Export Intensity than New Zealand's exporting industries. The graph on the right shows that Finnish production incorporates more salary and wages into value-added, possibly indicating that they produce more valuable, knowledge-intensive products – though it might mean something else, as well; all that is certain is that New Zealand tends to provide higher returns to capital than Finland (this is particularly interesting as New Zealand is considered within the OECD to be a low-wage economy, so perhaps it needn't be). New Zealand does have industries with relatively high wages as a proportion of value-added, but these industries have modest Export Intensity and are not as reliant on imports for intermediate inputs. There are only two New Zealand industries with very high Export Intensity and these are relatively reliant on domestic production for intermediate inputs, and the products



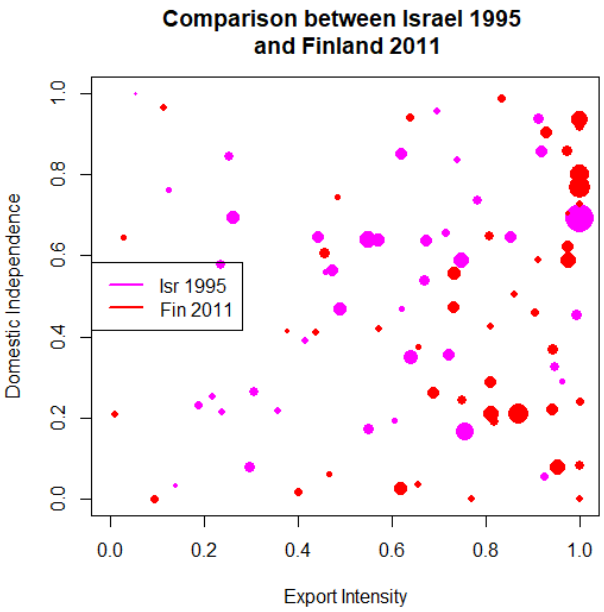
do not require highly paid workers to produce it or the workers are poorly paid relative to value-added; these industries are Education, and Food, Beverages, and Tobacco.

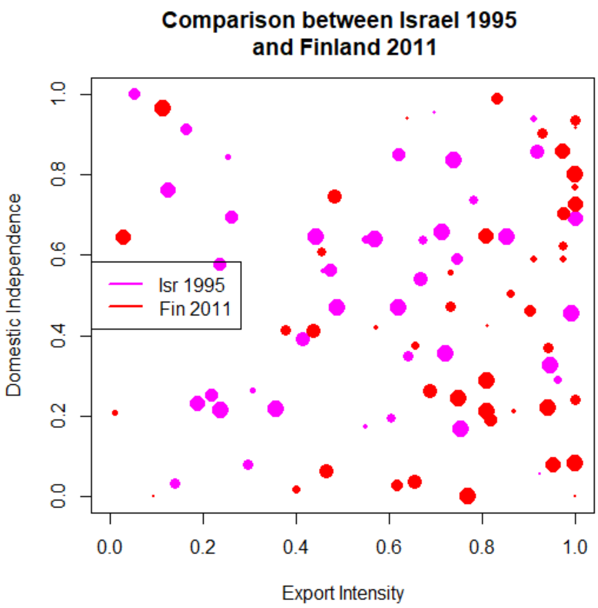
On the face of it, for New Zealand to become more like Finland would require an increase in the number of significantly exporting New Zealand industries in the top right-hand corners of the graphs. If this corner is defined as being the 80th percentile and above for both Export Intensity and Domestic Independence, then Finland has 6 industries to New Zealand's 1: Mining and Quarrying.

Which industries would be likely candidates? Is the best course of action to promote industries near to the top right-hand corner to simultaneously export more or export more valuable products and become more efficient in its use of domestic and imported inputs? Or should there be sequential focus on these activities, with different plans for different industries? By considering the path in Economy Space between New Zealand and Finland we can generate possible plans, and these plans can subsequently be more stringently tested, perhaps by including an historical understanding of what was affecting the economies of Canada in 2005, Portugal in 2005, and Israel in 1995. That analysis is out of the scope of this work, requiring a lengthier treatment, but we will generate some plans that could be evaluated.

We begin by working backwards, considering the difference between Israel in 1995 and Finland in 2011.

From Israel to Finland





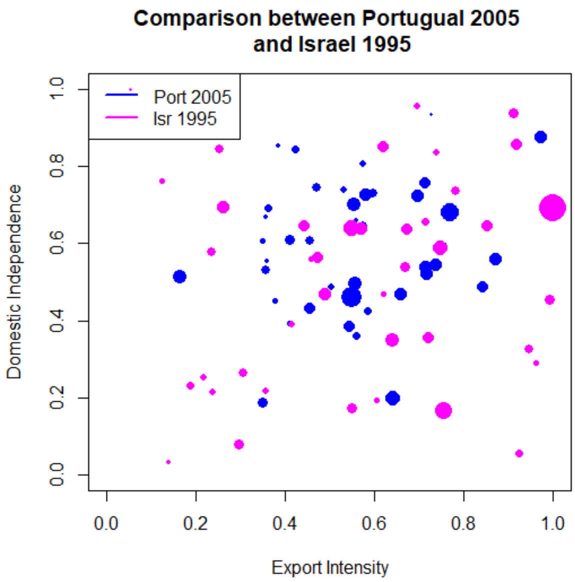
There are two noticeable differences between Israel and Finland:

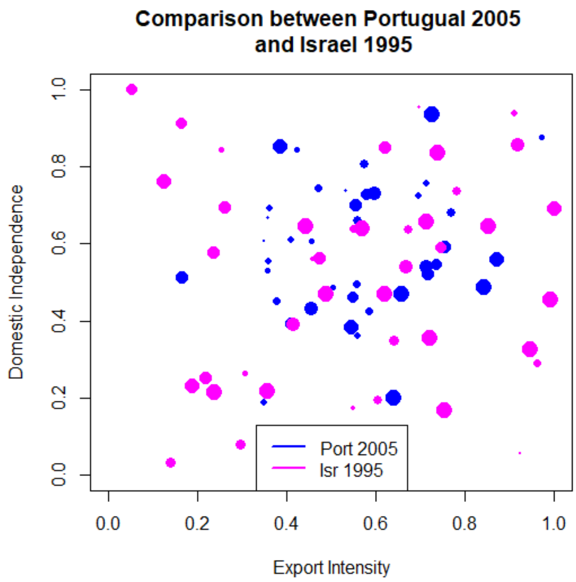
- Finland has more industries in the top right-hand corner
- For Finland, Export Intensity is strongly associated to the proportion of total exports, whereas Israel has many more significant exporters who have only middling Export Intensity.

The course of action to remove these differences is:

1. Israeli industries which have very high Export Intensity but not high Domestic Independence should take action to increase their Domestic Independence while maintaining their high Export Intensity.
2. The Israeli industries that have moderate Export Intensity and Domestic Independence but which are significant exporters should increase their Export Intensity, through increasing the volume or price of exports.

Portugal to Israel





The essential differences between Portugal and Israel are:

- The Israeli industries tend to have a larger proportion of their value-added in salary and wages.
- Portugal has very few industries with high Export Intensity.
- Israel has a number of low Export Intensity exporters of knowledge-intensive products, that have high Domestic Independence; Portugal does not.

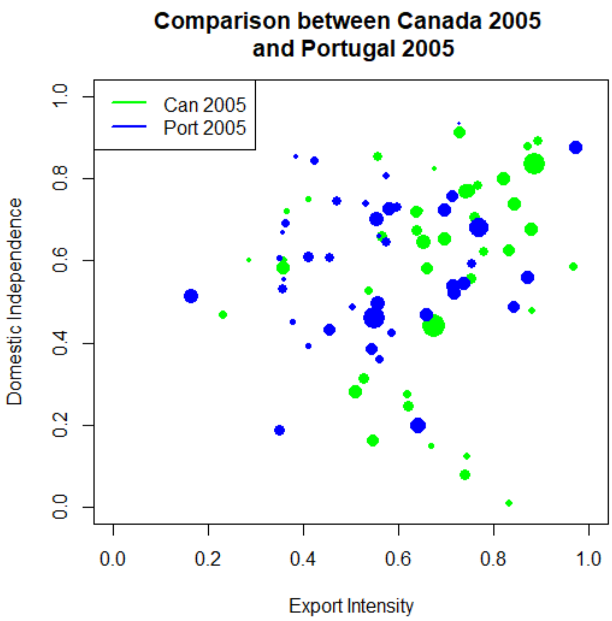
Thus, a possible course of action would be:

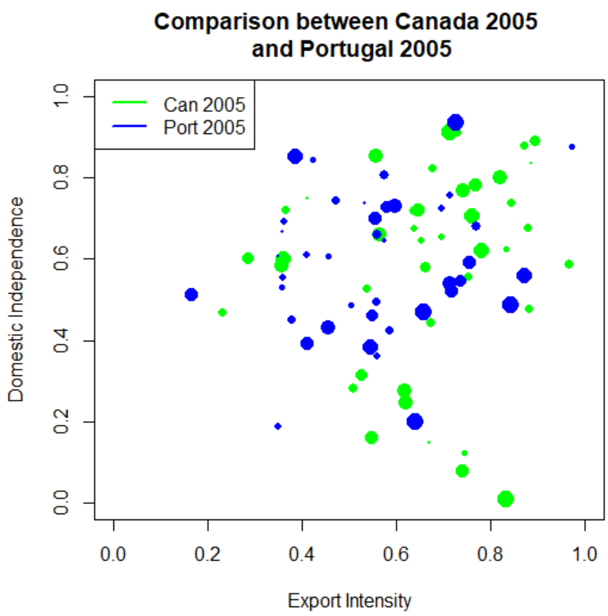
1. Encourage industries with moderate Export Intensity and Domestic Independence to increase their Export Intensity through an increase in the price of export products, creating products that are more knowledge intensive.
2. Encourage knowledge-intensive industries with moderate Export Intensity and high Domestic Independence to increase Export Intensity through an increase in export volumes.

3. Encourage knowledge-intensive industries with low Export Intensity and moderate Domestic Independence to increase their Domestic Independence through increased use of imports.

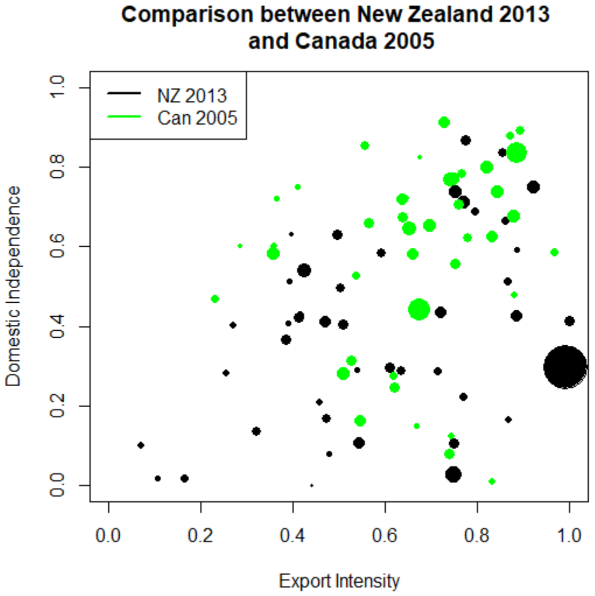
Canada to Portugal

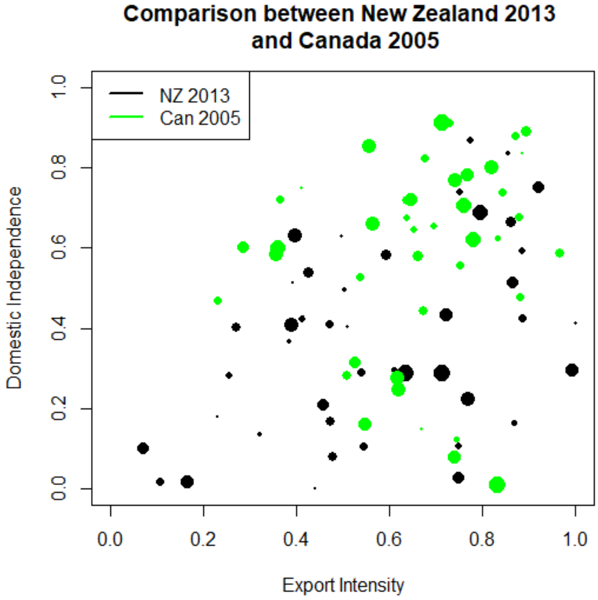
Canada and Portugal are very similar economies, and it appears that Canada is the more successful exporting economy, with more industries having both high Export Intensity and high Domestic Independence. Portugal is 'better' in that Portugal does not have the scattering of high Export Intensity, low Domestic Independence industries in the lower left-hand corners of the graph. Thus, these industries should be encouraged to increase their Domestic Independence.





New Zealand to Canada



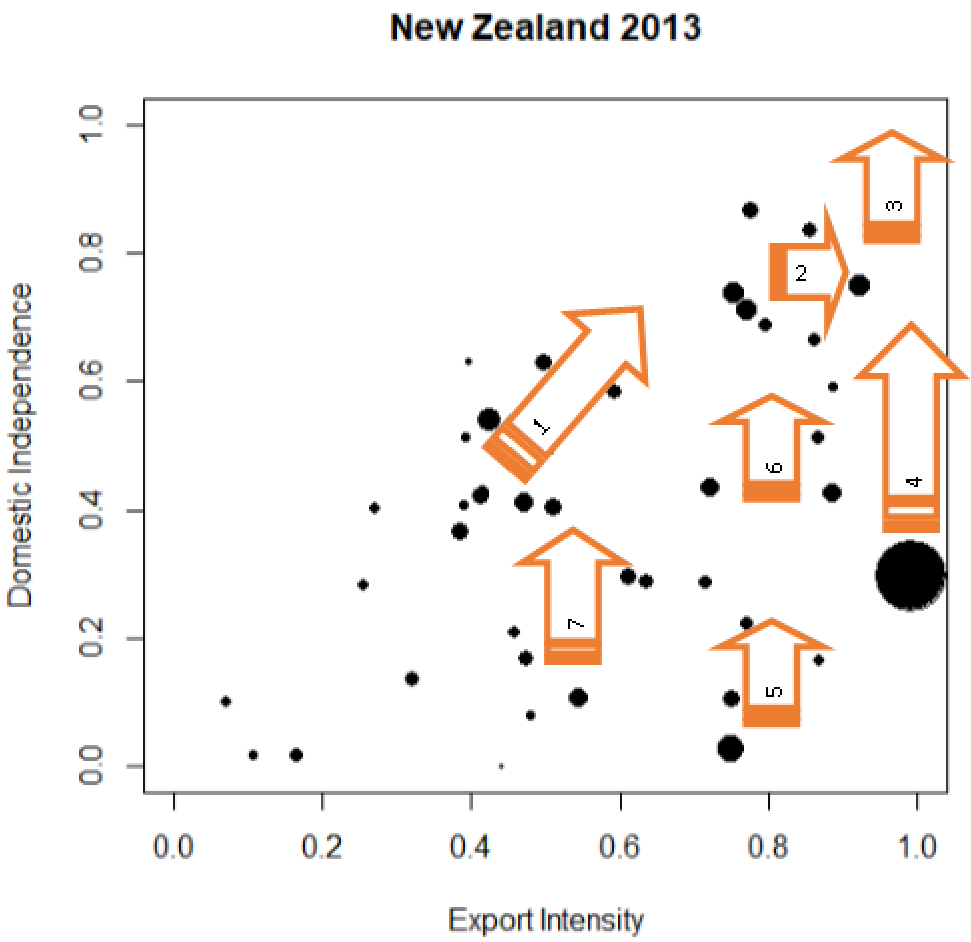


New Zealand and Canada appear to have a similar make-up of industries in the region of the graphs with high Export Intensity and high Domestic Independence. A key difference between the two economies is in the 'middle ground' of moderate Export Intensity and moderate Domestic Independence: New Zealand has a set of industries with both values around 0.5 that are moderate exporters of knowledge-intensive products, whereas the analogous group for Canada has greater Export Intensity and Domestic Independence, about 0.7. Also, Canada does not have any low Export Intensity industries.

Thus, the following action is suggested:

- Encourage the exporting industries with moderate Export Intensity and moderate Domestic Independence to increase both of these measures at the same time.
- Encourage high Export Intensity NZ industries with low Domestic Independence to increase their Domestic Independence while maintain high Export Intensity.

Graphically, we depict the strategy for increasing New Zealand's similarity to Finland below.



In the graph above, each arrow indicates the desired development of the industries at the tail of the arrow. That development might be through producing more valuable products, through exporting more, through using more imports for intermediate inputs, or through more efficient use of inputs.

Food, Beverages, and Tobacco is New Zealand's largest exporting industry, accounting for 39% of exports. We now consider the application of the strategy to it and those industries that have more than the average of what remains, about 1.3% - together with Food, Beverages, and Tobacco this gives 18 significant exporting industries accounting for 87% of exports.

We discuss these 18 industries in turn, at the risk of exhausting the reader's patience.

Food, Beverage, and Tobacco

Representing 39% of exports, this industry is in the 99th percentile for exports as a percentage of domestic value-added in production, compared to all 115 Food, Beverage, and Tobacco industries in the STAN set of input-output tables.

However, for Domestic Independence it is only in the 30th percentile, and for salaries and wages as a percentage of value-added it is in the 77th percentile – which is extremely high for New Zealand, where the median for all industries is the 49th percentile. Thus, the strategy would be to reduce the dependence on other NZ industries' inputs.

There may not be very much scope for this. Unsurprisingly, 53% of inputs in Food, Beverage, and Tobacco are from Agriculture, hunting, forestry and fishing, and this ability to source primary ingredients domestically is the basis for New Zealand's competitive advantage in Food, Beverage, and Tobacco exports.

Other industries contribute much smaller proportions of intermediate input, and they would seem to be such that their input is not substitutable by imports. These industries are Wholesale and retail trade (4%); Land transport, or transport via pipeline (4%); Other business activities (3%) – in NZ this is professional services (veterinary, accounting, legal, employment and labour hire, marketing and publishing); and Finance and insurance (2%). It is conceivable that there are efficiency gains that could be had, but this is unlikely to shift Domestic Independence very much.

We conclude that this analysis can only suggest that Food, Beverage and Tobacco seek to increase exports of high-value added products – as is the commonly agreed export strategy.

Agriculture, hunting, forestry and fishing

This industry is responsible for 7.5% of exports, and is in the 75th percentile for exports as a proportion of domestic value-added in production; it is in the 3rd percentile for Domestic Independence, and the 68th percentile for salary and wages as a proportion of value-added.

Given how much of the output of this industry is used as intermediate inputs in Food, Beverage, and Tobacco, it may not be feasible to significantly increase the proportion of production that is exports. The low Domestic Independence is due to some significant intermediate inputs: 8% of inputs by Wholesale and retail trade; 7% by Finance and Insurance; 5% by Real estate activities; 4% each by Chemicals (excluding pharmaceuticals), Other business activities, and Land transport.

The high levels of input by Finance and Insurance and by Real estate activities are remarkable, and it would be interesting to investigate whether the low levels of Domestic Independence are a result of these costs.

The only obvious set of tradeable inputs are those from the Chemicals industry, which is in the 50th percentile for exports as a proportion of domestic value-added amongst all Chemical industries in the data set, but which has relatively small amounts of export.

If Agriculture, hunting, forestry and fishing were to increase its Domestic Independence by importing its Chemicals industry input, it would be sensible to support the NZ Chemicals industry in developing overseas markets.

Wholesale and retail trade; repairs, Hotels and restaurants, and Air Transport

Exports from these three industries (which provide 5%, 4% and 4% of exports, respectively) constitute the Tourism sector, including goods and services to visiting sea-craft and aircraft.

The following table contains the benchmark values:

Industry	Domestic Independence	Salary and wages share of value-added	Export Intensity
Wholesale and retail trade; repairs	54th percentile	62nd percentile	43rd percentile
Hotels and restaurants	75th percentile	68th percentile	92nd percentile
Air transport	74th percentile	42nd percentile	75th percentile

The strategy indicates that Wholesale and retail trade should both reduce its domestically sourced inputs and increase its exports (either in value or in volume). In terms of substituting domestic inputs for imports, the industries that are significant providers of intermediate inputs to Wholesale and retail trade are: Other business activities (14%); Real estate activities (12%); Finance and insurance (11%); Transport support activities (8%); and Land transport (7%).

The strategy for Hotels and restaurants would be to increase Domestic Independence, even though it is currently high for Domestic Independence. Key industries providing intermediate inputs are Food, Beverages, and Tobacco (24%); Real estate activities (11%); Wholesale and retail trade (8%); Other business activities (7%); and Finance and insurance (6%). So, again we see that there are large inputs by the professional services industries (including marketing and employment services), financial and insurance services, and real estate activities.

The strategy for Air Transport would be to increase exports, which presumably means an increase in tourist numbers or an increase in New Zealander's travelling abroad. Increasing tourist numbers can be done by having shorter stays but more frequent visits, or by having more visitors; more visitors would require that the Hotel and restaurant infrastructure be able to cope with more, which may have implications for investment. We also note that the salary and wages costs for Air Transport are below the median, so that presumably the airlines would be able (as a whole) to weather some of the costs of expanding the number of flights before the demand might truly warrant, perhaps at the risk of reduced dividends to shareholders.

We raise the question of whether the cost of professional services and marketing can be reduced for these industries, or whether rental costs or finance costs are having a detrimental impact on the ability of the Tourism sector to be competitive, or making the New Zealand economy more susceptible to any downturn in tourism.

Mining and quarrying (energy)

This industry category in New Zealand consists of Coal mining, Oil and gas extraction, and Exploration and other mining support services, and accounts for 4% of exports. It is in the 77th percentile for Export Intensity and the 71st percentile for Domestic Independence, so the strategy suggests it increase both.

The key inputs into Mining and quarrying (energy) are Research and development (13%); Finance and insurance (7%); Land transport, including pipelines (5%); and Other business activities (4%). In New Zealand, the Research and development industry is synonymous with Scientific, architectural, and engineering services. Encouraging a shift from a reliance on domestic Research and development inputs to imported inputs could be done by facilitating the export of Research and development services.

In line with the other industries we have looked at, it would be worth investigating whether the cost of Finance and insurance inputs reduces competitiveness and productivity.

The industry is in the 28th percentile for salary and wages in proportion to value-added, so that the return on capital is quite high in comparison to other Mining and quarrying (energy) sectors world-wide.

Office, accounting and computing machinery

For NZ, this industry consists of Electronic and electric equipment manufacturing, and accounts for 2% of exports. It is in the 41st percentile for Domestic Independence and the 47th percentile for Export Intensity; it is also in the 47th percentile for salary and wages in proportion to value-added.

The strategy indicates that we should encourage an increase in both Domestic Independence and Export Intensity at the same time. Thus, more and higher-value exports while increasing efficiency in use of inputs or substituting domestic inputs for imports.

The industry already obtains 42% of its inputs from overseas, but there are industries providing a large amount of domestic inputs: 10% of inputs comes from Wholesale and retail trade, and 7% from Other business activities. It is possible that the large percentage of input from Wholesale and retail trade is due to the large amount of domestic consumption of output from the Office, accounting and computing machinery industry, so that by increasing exports that percentage will decrease and both Domestic Independence and Export Intensity will increase. Or it could be that the industry needs assistance in directly sourcing its inputs from overseas, bypassing the need for middlemen. We note also, that the low benchmark for salaries and wages as a proportion of value-added could indicate that the industry will have difficulty in developing high value-added products for export as talent seeks better compensation elsewhere, though equally there is opportunity to increase compensation should that be required.

Manufacturing nec; recycling (include Furniture)

This industry includes Pulp, paper, and converted paper manufacturing; Furniture manufacturing; Other manufacturing; and Waste collection, treatment and disposal services, accounting for 2% of exports. The industry is in the 11th percentile for Domestic Independence, the 54th percentile for Export Intensity, and the 50th percentile for salaries and wages in proportion to value-added.

The strategy applied to the industry indicates an increase in Domestic Independence, which means an increased use in imported intermediate inputs. Many of these inputs are not readily importable or would be nonsensical to import given New Zealand's factor advantages: Production, collection and distribution of electricity (7% of inputs); Agriculture, hunting, forestry and fishing (7%); Wood and products of wood and cork (7%); and Land transport (4%). Certainly, there may be efficiencies to be gained, but substitution is unlikely to be a viable strategy.

The other industries providing significant inputs are Wholesale and retail trade (9%); and Fabricated metal products (4%). Small increases to Domestic Independence might be gained through more direct supply-chain management and a shift to importing fabricated metal products, with a commensurate redirection of fabricated metal products to exports.

Wood and products of wood and cork

Accounting for 2% of exports, this industry has Domestic Independence in the 43rd percentile, Export Intensity at the 72nd percentile, and salary and wages in proportion to value-added in the 75th percentile. The strategy indicates that the industry should increase its Domestic Independence.

Key input industries are Wholesale and retail trade (10%); Other business activities (6%); Office, accounting and computing machinery (6%); Finance and insurance (5%); and Real estate activities (4%). As in other industries, we wonder whether the industry can more directly manage its supply chains, reducing its need for wholesale and retail trade inputs.

Other community, social and personal services AKA 'Film, television, and music production'

Also 2% of exports, this industry has Domestic Independence in the 42nd percentile, Export Intensity in the 88th percentile, and salary and wages in proportion to value-added in the 50th percentile. A 'grab-bag' of industries, in New Zealand this is dominated by Motion picture and sound recording activities.

The strategy would be to increase Domestic Independence. The industries providing intermediate inputs are Other business activities (14%); Finance and insurance (8%); and Real estate activities (7%). Arguably it is either infeasible or nonsensical to substitute these inputs with imports; it is worth investigating whether the cost of these inputs is a competitive disadvantage to the industry increasing its exports.

Education

The Education sectors (Primary, Secondary and Tertiary) are responsible for 2% of exports. The Domestic Independence is in the 41st percentile, Export Intensity is in the 100th percentile, and salary and wages in proportion to value-added is in the 20th percentile.

These figures deserve some examination. Given that New Zealand Education is very export intensive and most other economies Education industries are to a much lesser degree, it is perhaps reasonable that salary and wages as a proportion of value-added is so relatively low; overseas students will pay a premium over domestic students, thus increasing the amount of value-added.

Most Education industries will not be able to import intermediate inputs, if their inputs are similar to what the NZ Education industry uses: Other business activities (9%); Other community, social, and professional services (9%); Wholesale and retail trade (6%); Construction (6%); Finance and insurance (6%); Electricity (5%); and Real estate activities (5%). Hence, we hypothesise that Education industry in New Zealand gets its low Domestic Independence score because these inputs are more expensive than is usual or the industry is inefficient and could use less.

Textiles, textile products, leather and footwear

This industry accounts for 2% of exports and is in the 10th percentile for Domestic Independence, 38th percentile for salary and wages in proportion to value-added, and the 75th percentile for Export Intensity. The strategy indicates that the industry should seek to increase its Domestic Independence.

Once again, this is a situation where the key inputs are sources of natural advantage to New Zealand, namely Agriculture, hunting, forestry and fishing (22%), and Food, Beverages and Tobacco (6%). The other key inputs are Wholesale and retail trade (11%) and Other business activities (5%); given how export focused the industry is, this cost to Wholesale and retail trade seems worth further investigation.

We note that as production by Agriculture, hunting, forestry and fishing reaches its maximum, this industry will need to source inputs from overseas if it is to increase export volumes – this will increase both Domestic Independence and Export Intensity.

Should the industry seek to increase Export Intensity by producing more valuable products, there might be cause for concern that the industry will be unable to attract skilled labour under existing compensation levels.

Machinery and equipment, nec

As 2% of exports, this industry has Domestic Involvement in the 40^{th} percentile (with 33% of inputs from imports), Export Intensity in the 51^{st} percentile, and salary and wages in proportion to value-added in the 21^{st} percentile.

The strategy would be to increase both Export Intensity and Domestic Involvement. Domestic inputs are possibly higher than the industry norm because key inputs such as Iron and steel (12% of inputs) are available domestically (often that is not the case). Other key inputs are Wholesale and retail trade (9%); non-Ferrous metals (7%); and Fabricated metal products (7%). Areas to investigate are whether the industry could manage its supply chain internally, without relying on Wholesale and retail trade. Increased exports in this industry are likely to be of high value-added products. The low rating for salary and wages could indicate that the industry is not investing in the research and development needed to develop these products.

Supporting & auxiliary transport activities; activities of travel agencies

Part of this industry's exports belongs to the Tourism industry, but it also includes transport logistics and warehousing so we separate it from the Tourism industries above. As 2% of exports, this industry is in the 42^{nd} percentile for Domestic Independence, the 45^{th} percentile for salary and wages in proportion to value-added, and the 41^{st} percentile for Export Intensity.

The strategy would be to increase both Export Intensity and Domestic Independence. Exports of logistics capability is a service industry, so the challenge might be to increase staff compensation to stay competitive – there is room to do so. Other production by this industry scales with the amount of exports.

In terms of increasing Domestic Independence, the two largest domestic inputs are Other business activities (15%) and Real estate activities (14%) – additional investigation is required to understand the nature of these costs.

Computer and related activities

In New Zealand, this industry consists of Broadcasting and internet publishing, and Computer system design and related services (software development). It accounts for 2% of exports, it is in the 63rd percentile for Domestic Independence, the 50th percentile for Export Intensity, and the 23rd percentile for salaries and wages as a percentage of value-added.

The mix of software design vs internet publishing may be different in other countries, so the comparisons may not be entirely straightforward. Nevertheless, the strategy would be to increase both Export Intensity and Domestic Independence.

Increased exports would require investment in research and development or having skilled staff; the low ranking for salaries and wages implies that there is an ability to spend more there that is not being utilised.

The biggest domestic inputs are from Finance and insurance (11%), Other business activities (9%), Real estate (7%), and Wholesale and retail trade (6%). These are not readily substitutable by imports, but perhaps efficiencies are possible. The high cost for Finance and insurance is worth investigating.

Finance and insurance

This industry is around 2% of exports, is in the 30th percentile for Domestic Independence, the 61st percentile for Export Intensity, and the 31st percentile for salary and wages as a proportion of value-added. Once again, a service industry that appears to have very high returns to capital.

The strategy is to increase both Export Intensity and Domestic Independence. Key inputs are Other business activities (11%) and Real estate activities (8%); given the low ranking for Domestic Involvement, presumably internationally these costs are either lower or not incurred because the industry supplies the inputs ‘in-house’. Thus, the low Domestic Independence may be due to the industry’s size and inability to maintain continual access to the skills it needs – but that is speculation, and more investigation is warranted.

Other Business Activities

This industry covers a wide variety of professional services: legal service, accounting, employment services, marketing services, plus publishing. It accounts for 2% of exports, is in the 37th percentile for Domestic Involvement, the 38th percentile for Export Intensity, and the 31st percentile for salary and wages as a proportion of value-added – again, odd for a services industry, but perhaps the capital costs of publishing skews the comparison.

The strategy is to increase Domestic Independence. The key inputs into this industry are Pulp, paper, paper products (11%), Finance and insurance (9%), Real estate activities (7%) and Supporting and auxiliary transport activities (6%). Given that New Zealand is a producer of paper products, it doesn’t seem viable to substitute these inputs with imports.

Further investigation is required to understand how this industry can increase its Domestic Independence and Export Intensity.

Coke, refined petroleum products and nuclear fuel

Our final of the 18 industries, this industry performs about 2% of exports, has Domestic Independence in the 87th percentile, Export Intensity in the 77th percentile, and salary and wages as a proportion of value-added in the 42nd percentile.

We should increase the exports by this industry, which will come about from activity by the Mining and quarrying (energy) industry, subsequently increasing that industry's ability to export in the Exploration and other mining support services.

Concluding the case study

In conclusion, there is opportunity to bring the New Zealand economy closer to that of Finland in 2011, and progressing New Zealand towards its Business Growth Agenda target of 40% exports to GDP. The analysis in the preceding pages can be refined and extended to define a set of changes to the country's industries, and then policies crafted to enable, effect, or encourage these changes. Such policies could involve directing trade promotion activity, creating instruments whereby the government can absorb particular risks or costs to facilitate aspects of the economy, set immigration or education policy, et cetera. Economy Space analysis can be used in evaluating how well these policies, and the strategy they support worked.

Many exporting industries could export more or export higher-value products, such as Electronic and electrical equipment, Fabricated metal products, Mining and quarrying (energy), Chemicals, Scientific, architectural, and engineering, services, and Computer system design, amongst others; there are generally low wages in these industries as a proportion of value-added, which could indicate a lack of investment in the research and development needed to create high-value products. Generally, these industries are quite able to fund such activities – as evidenced by the proportion of value-added that is salaries and wages.

Some industries have high export intensity and if they are to increase exports, it is likely that investment will be required. These industries are Tourism and Education.

Other industries should decrease their reliance on New Zealand-sourced inputs or use their inputs more efficiently. It would be worth investigating how these industries can be supported in more efficiently managing their supply chains, leading to a reduction in wholesaling and retailing expenses. Other significant expenses are Finance and insurance, and Real estate activities – an understanding of whether these expenses are particularly high for New Zealand industries and, if so, the impact on the economy is needed.

Two industries largely dominate New Zealand exports: Food, Beverages, and Tobaco, and Agriculture, hunting, forestry, and fishing. There is relatively little scope for these industries to reduce their domestic inputs, and are already highly export intensive. Until these industries significantly shift towards high value-added exports, or other industries reach significant export values, there is little hope that New Zealand will reach its Business Growth Agenda goals.

References

Bergstrom, M. R. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences, 105*(4), 1118-1123.

Blondel, V. D., Guillaume, J.-L., & Lambiotte, R. &. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 10*, P10008.

Bornholdt, J. R. (2006). Statistical mechanics of community detection. *Physical Review E, 74*(1).

Gabor Csardi and Tamas Nepusz. InterJournal, C. S. (2006). The igraph software package for complex network research. *InterJournal Complex Systems, 1695*.

Latapy., P. P. (2006). Computing communities in large networks using random walks. *J. Graph Algorithms Appl., 10*(2), 191-218.

Ministry of Business, Innovation, and Employment. (2017, 2 21). *Business Growth Agenda*. Retrieved from MBIE: http://www.mbie.govt.nz/info-services/business/business-growth-agenda

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences, 103*(23), 8577-8696.

OECD. (2016). *STAN database*. Retrieved from OECD: http://www.oecd.org/sti/stan

OECD and WTO. (2012). *Trade in value-added: Concepts, methodologies and challenges.* OECD. Retrieved from www.oecd.org/sti/ind/49894138.pdf

R Core Team. (2015). R: A Language and Environment for Statistical Computing. . Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Statistics Finland. (2017). Official Statistics of Finland (OSF): Input-output. Helsinki. Retrieved from http://www.stat.fi/til/pt/tau_en.html

Statistics New Zealand. (2012). Statistics New Zealand. National accounts input-output tables: year ended march 2007. Retrieved from http://www.stats.govt.nz/~/media/Statistics/browse-categories/economic-indicators/national-accounts/supply-use-tables/input-output-tables-ye-ended-mar-07.xls

Statistics New Zealand. (2016). National accounts input-output tables for the year ending March 31st 2013. Retrieved from http://www.stats.govt.nz/~/media/Statistics/browse-categories/economic-indicators/national-accounts/supply-use-tables/na-input-output-tables-ye-mar13.xlsx